REMARKS

Claims 1, 2 and 4-9 are pending in the application. By this Amendment, claim 1 is amended. The amendments to claim 1 are solely to correct errors in recitation of terms in order to provide proper antecedent basis. No new matter is added.

Reconsideration of the application is respectfully requested in view of the above amendments to the claims and the following remarks. For the Examiner's convenience, Applicant's remarks are presented in the order in which they were raised in the Office Action.

Rejections Under 35 U.S.C. § 112

(i) Clams 1-2, and 4-9 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement.

The Examiner states that the specification as filed does not provide a sufficient description of the Egr-1 gene sequence or the Egr-1 gene products that are encompassed by the instant claims.

Applicants respectfully traverse. Applicants submit that prior to the filing date of the instant application (January 9, 2001) and even prior to the filing date of the priority Australian Application PN8554 (filed March 7, 1996), the gene (nucleic acid) and protein sequence of Egr-1 was publicly available to one of skill in the art.

Applicants refer the Examiner to page 10, lines 16-21 of the Specification of the instant application citing an article entitled "Early Growth Response Protein 1 (Egr-1): Prototype of a Zinc-

finger Family of Transcription Factors" (Gashler A and Sukhatme VP, *Prog. Nucleic Acid Res. Mol. Biol.* 50:191-224 (1995). The article in its entirety is incorporated by reference into the Specification as specified on page 21, lines 25-26 therein. Figure 3 of the Gashler and Sukhatme reference (p. 204) shows the schematic structure and amino-acid sequence of the Egr-1 protein.

Further, prior to the filing and priority date of this application Egr-1 gene and amino acid sequences from a wide variety of species (human, mouse, chicken and zebrafish) had been identified and characterized (*see* page 203 of Gashler and Sukhatme) and found to be "highly homologous." Characteristic structural attributes of Egr-1 proteins were also well recognized and Egr-1 proteins were known to contain three CysCys—HisHis zinc-finger motifs. Page 203 of Gashler and Sukhatme discuss these attributes.

Thus, the Specification and references incorporated into the Specification provides not only functional characteristics of Egr-1 but structural (sequence) description of Egr-1 in satisfaction of the Written Description Guidelines and MPEP 2163.

Applicants also include herewith additional references (published prior to the priority date of this application) which disclose the gene and amino acid sequences of Egr-1 and were available to one of ordinary skill in the art who would thus have concluded that the inventors had possession of the claimed invention. Figure 1 of Joseph *et al.* shows a comparison of the sequences of Egr-1 and Egr-2 ("Molecular cloning, sequencing, and mapping of EGR2, a human early growth response gene encoding a protein with "zinc-binding finger" structure" *Proc Natl Acad Sci U S A.* 85(19):7164-7168 (1988 Oct)). Suggs *et al.* provide a complete cDNA sequence of Egr-1. (Suggs SV, Katzowitz JL, Tsai-Morris C, Sukhatme VP. "cDNA sequence of the human cellular early

growth response gene Egr-1" *Nucleic Acids Res.*18(14):4283 (1990 Jul 25)). The abstracts of two other published articles which provide the gene and amino acid sequences of Egr-1 are also included for the Examiner's reference: (i) Sukhatme *et al.* " A novel early growth response gene rapidly induced by fibroblast, epithelial cell and lymphocyte mitogens" *Oncogene Res.* 1987 Sep-Oct;1(4):343-355; and (ii) Sukhatme *et al.* " A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization" Cell. 1988 Apr 8;53(1):37-43.

Applicants submit that the Specification, references incorporated therein and the state of the art prior to the filing date of Applicants' priority application provided sufficient guidance to one of skill in the art about the structural and functional characteristics of Egr-1 to conclude that the Applicants were in possession of the claimed invention at the time of filing. Therefore, applicants respectfully request that the "written description" rejection under 35 USC §112, first paragraph be withdrawn.

(ii) Claims 1-2 and 4-9 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 1 stand rejecetd for indefiniteness over lack of antecedent basis for the term "putative compound."

In response, applicants amend claim 1(a), line 4 to recite "**a** putative compound" (emphasis added) and amend claim 1(b), line 8 to "**the** putative compound" (emphasis added) in order to provide proper antecedent basis for this term in the claim.

In addition, claim 1(a), line 4 is amended to replace "the" ability with "an" ability for the purpose of removing any ambiguity regarding antecedent basis for the term "ability.".

In view of these amendments to claim 1, Applicants believe that the terms recited in claim 1, as amended, have proper antecedent basis and respectfully request withdrawal of this ground for rejection under 35 USC §112, ¶2. Claims 2 and 4-9 depend from claim 1.

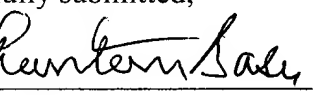
CONCLUSION

In view of the amendments and arguments set forth above, Applicants earnestly believe that they are entitled to a letters patent and respectfully request the Examiner to expedite prosecution of this patent application to issuance. Should the Examiner have any questions, the Examiner is invited to telephone the undersigned at the number given below.

In the event the U.S. Patent and Trademark office determines that an extension and/or other relief is required, applicant petitions for any required relief including extensions of time and authorizes the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to Deposit Account No. 03-1952 referencing docket no.529282000220. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: September 3, 2004

Respectfully submitted,

By 

Shantanu Basu
Registration No.: 43,318
MORRISON & FOERSTER LLP
755 Page Mill Road
Palo Alto, California 94304
(650) 813-5995

24. D. Kardassis, M. Hadzopoulou-Cladaras, D. P. Reczji, R. Cortese, V. I. Zannis and C. Cladaras, *MCBiol.* 10, 2553 (1990).
25. J. A. A. Ladias, M. Hadzopoulou-Cladaras, D. Kardassis, P. Cardot, J. Cheng, V. Zannis and C. Cladaras, *JBC* 267, 15849 (1992).
26. S. Metzger, T. Lell and J. L. Breslow, *JBC* 263, 9978 (1990).
27. E. H. Ludwig, B. Levy-Wilson, T. Knott, V. D. Blackhart and B. J. McCarthy, *Cell Biol.* 10, 329 (1991).
28. A. R. Brooks, and B. Levy-Wilson, *MCBiol.* 12, 1134 (1992).
29. B. Levy-Wilson, B. Paulweber, B. P. Nagy, E. H. Ludwig and A. R. Brooks, *JBC* 267, 18735 (1992).
30. B. Paulweber, A. R. Brooks, B. P. Nagy and B. Levy-Wilson, *JBC* 266, 21956 (1991).
31. B. Paulweber and B. Levy-Wilson, *JBC* 265, 24161 (1991).
32. B. D. Blackhart, Z. Yao and B. J. McCarthy, *JBC* 265, 8358 (1990).
33. D. B. Goring, J. Rossant, S. Clapoff, M. L. Breitman and L. C. Tsui, *Science* 235, 456 (1987).
34. A. R. Brooks, B. P. Nagy, S. Taylor, W. S. Simonet, J. M. Taylor and B. Levy-Wilson, *MCBiol.* 14 (1994). In press.
35. R. A. McKnight, A. Shamay, L. Sankaran, R. J. Will and L. Hennighausen, *PNAS* 89, 6943 (1992).
36. S. L. Hofmann, D. W. Russell, M. S. Brown, J. L. Goldstein and R. E. Hammer, *Science* 239, 1277 (1988).
37. X. C. Jiang, L. B. Agellon, A. Walsh, J. L. Breslow and A. Tall, *J. Clin. Invest.* 90, 1290 (1992).
38. F. Grosveld, G. B. van Assendelft, D. R. Greaves and G. Kollias, *Cell* 51, 975 (1987).
39. A. Schedl, L. Montoliu, G. Kelsey and G. Schütz, *Nature* 362, 258 (1993).
40. P. Fraser, S. Pruzina, M. Antoniou and F. Grosveld, *Genes Dev.* 7, 106 (1993).
41. W. S. Simonet, N. Bucay, S. J. Lauer and J. M. Taylor, *JBC* 268, 8221 (1993).
42. C. A. Pinkert, D. M. Ornitz, R. L. Brinster and R. D. Palmiter, *Genes Dev.* 1, 268 (1987).
43. B. J. Aronow, R. N. Silbiger, M. R. Dusing, J. L. Stock, K. L. Yager, S. S. Potter, J. J. Hutton and D. A. Wiginton, *MCBiol.* 13, 4170 (1993).
44. J. Mirkovitch, M.-E. Mirault and U. K. Laemmli, *Cell* 39, 223 (1984).
45. P. N. Cockerill and W. T. Garrard, *Cell* 44, 273 (1986).
46. C. Bonifer, M. Vidal, F. Grosveld and A. E. Sippel, *EMBO J.* 9, 2843 (1990).
47. V. C. Blasquez, M. Xu, S. C. Moses and W. T. Garrard, *JBC* 264 21183 (1989).
48. L. Emorine, M. Kuehl, L. Weir, P. Leder and E. E. Max, *Nature* 304, 447 (1983).
49. T. G. Parslow and D. K. Granner, *Nature* 299, 449 (1982).
50. T. G. Parslow and D. K. Granner, *NARes.* 11, 4775 (1983).
51. M. P. Lisitsa, R. V. Farese, Jr., G. Chiou, D. S. Goss, F. Chin, R. E. Hammer, H. H. Hobbs and S. C. Young, *J. Clin. Invest.* 92, 3029 (1993).
52. M. J. Callow, L. J. Stoltzfus, R. M. Lawn and E. M. Rubin, *PNAS* 91, 2130 (1994).

Early Growth Response Protein 1 (Egr-1): Prototype of a Zinc-finger Family of Transcription Factors

ANDREA GASHLER[1] AND VIKAS P. SUKHATME[2]

Department of Medicine
Beth Israel Hospital and Harvard Medical School
Boston, Massachusetts 02215

I. Overview of Immediate-early Genes

Extracellular signals in the form of soluble factors, matrix proteins, and adhesion molecules influence the proliferation and differentiation of eukaryotic cells. These long-term responses, mediated by changes in gene expression, are coupled to biochemical events occurring in the plasma mem-

[1] Present address: Ligand Pharmaceuticals, San Diego, CA 92121.
[2] To whom correspondence may be addressed.

brane and cytosol that follow ligand–receptor interactions or other changes in the extracellular milieu. The so-called immediate-early genes are the earliest downstream nuclear targets for these events. These genes are, by definition, induced in the absence of *de novo* protein synthesis. In particular, a subclass of these genes encodes transcription factors, and these products form the first step in a cascade of gene–protein interactions. Thus, immediate-early transcription factor genes serve as nuclear couplers of early cytoplasmic events to long-term alterations in gene expression.

At present, the best characterized members of this group include c-*fos*, c-*jun*, and *Egr-1*. In turn, each of these genes is a prototype for a family of closely related proteins. This review focuses on the *Egr* gene family and its most extensively characterized member, *Egr-1*, first identified as an immediate-early gene responsive to growth factors and various differentiation cues, later confirmed to be a transcriptional regulatory protein. Other reviews have focused on changes in gene expression during the cell cycle (*1*) and transcriptional responses to extracellular signals (*2–4*). As a group, immediate-early transcription factors have provided important insights into how cellular responses to diverse extracellular signals are mediated.

II. Identification of *Egr-1* cDNA by Differential Screening

One approach to identifying novel genes that play key roles in cellular growth control is to focus on transcripts whose expression is low in nondividing cells but is rapidly up-regulated in cells stimulated by mitogen. Using c-*fos* as a model of immediate-early gene induction, several groups used similar differential screening strategies to isolate novel genes induced without intervening protein synthesis. Specifically, the following criteria were applied in our screen for important regulators of the G_0–G_1 transition: (1) Transcripts should be induced rapidly by serum stimulation of quiescent fibroblasts; (2) the mitogenic induction should not be affected by inhibitors of protein synthesis, such as cycloheximide; (3) expression should be induced by a spectrum of mitogens in a wide variety of cell types; and (4) the genes should be highly conserved in evolution (*5, 6*). In particular, we pursued differential screening of a library from BALB/c 3T3 cells stimulated for 3 hours with serum in the presence of cycloheximide. Clones were identified that hybridized preferentially to cDNA from serum and cycloheximide-treated fibroblasts as compared to cDNA from quiescent cells. The immediate-early gene c-*fos* was reisolated by this protocol. In addition, mitogenic stimulation of a variety of cell types from different species induced a 3.4-kb transcript. This novel immediate-early gene, designated *Egr-1* (*5–7*), has been independently cloned by similar differential screening strategies by a number of groups: *NGFI-A* was isolated as a nerve growth factor-inducible transcript in rat pheochromocytoma PC12 cells (*8*); *zif268* was cloned from serum-stimulated BALB/c 3T3 fibroblasts (*9*); *tis8* was identified as a phorbol-inducible gene in 3T3 cells (*10*); the chicken homolog, *cef5*, was cloned as a v-*src*-inducible gene from chicken embryo fibroblasts (*11*); and gene *225* was identified as a T-cell-activated transcript (*12*). Through hybridization to a highly conserved domain of the *Drosophila* factor Krüppel, *Krox24* was isolated from serum-stimulated 3T3 cells (*13*).

III. *Egr-1* Is Expressed in Response to Diverse Stimuli

A. Induction by Mitogens

In response to mitogens such as growth factors, hormones, and the tumor promoter TPA (phorbol), *Egr-1* induction is universal. In addition, *Egr-1* is expressed in diverse physiological contexts in particular cell types. The broad spectrum of extracellular stimuli that induce *Egr-1* can be roughly subgrouped into four categories: (1) mitogens, (2) developmental or differentiation cues, (3) tissue or radiation injury, and (4) signals that cause neuronal excitation.

In every cell type examined, *Egr-1* expression is rapidly induced by mitogenic stimulation. For example, in quiescent 3T3 cells stimulated with fetal calf serum, *Egr-1* expression is seen as early as 10 minutes, peaks around 30 minutes, and decays rapidly thereafter, returning to basal levels by 3–4 hours. Purified growth factors such as platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and epidermal growth factor (EGF) also stimulate *Egr-1* expression in fibroblasts (*5, 9, 13*). The kinetics of induction are similar to those of c-*fos*, but the magnitude of *Egr-1* induction is typically severalfold greater (*5*).

In addition to induction in fibroblasts, mitogenic stimulation of *Egr-1* has been described in a wide array of cell types, such as kidney and liver epithelial cells and lymphocytes. For example, *Egr-1* is induced in regenerating liver within 1 hour after partial hepatectomy (*14*); in serum-starved BSC-1 monkey kidney epithelial cells in response to the mitogen adenosine diphosphate, in serum-deprived rat hepatoma H35 cells stimulated with serum or insulin, and in human peripheral blood lymphocytes treated with phytohemagglutinin (*5*). *Egr-1* is also up-regulated by protein tyrosine kinases, whose activity is associated with transformation in culture and tumorigenesis in animals. *Egr-1* message levels increase when a temperature-sensitive variant of v-Src is shifted from the nonpermissive to the permissive

temperature. *Egr-1* is similarly induced by expression of a second tyrosine kinase, v-Fos (*15, 16*). Because protein-tyrosine kinase activity has been implicated in events promoting cell division, *Egr-1* may be an important component of the mitogenic signal.

An extremely tight correlation between *Egr-1* expression and B lymphocyte activation has been established (*17*). B lymphocytes express surface immunoglobulin that acts as receptor for antigen. While mature B cells are activated by cross-linking surface immunoglobulin with anti-μ antibodies and respond by proliferating, immature B-cells, such as the WEHI-231 cell line, respond to anti-μ by down-regulation of proliferation and eventually cell death. The *Egr-1* response in mature and immature B lymphocytes differs accordingly: *Egr-1* is rapidly and transiently induced in mature B cells cross-linked with anti-μ but not in WEHI-231 cells treated identically. However, *Egr-1* can be induced to respond in WEHI-231 cells exposed to lipopolysaccharide (LPS), a treatment that protects these cells from the antiproliferative effects of anti-μ (*17*). The participation of *Egr-1* in positive versus negative signaling through surface immunoglobulin may be mediated by differential methylation of the gene. *Egr-1* is hypermethylated in immature B cells and in the WEHI-231 line. When an *Egr-1* reporter is transfected (*18*) into the WEHI line, it can be activated by anti-μ in contrast to the endogenous gene. Most convincingly, endogenous *Egr-1* can be induced in WEHI-231 cells treated with the inhibitor of methylation, 5′-azacytidine (*18*).

Additional correlation of *Egr-1* induction with mitogenicity has been shown in studies (*19*) in rat kidney mesangial cells. Numerous vasoactive agents, including PDGF, vasopressin, serotonin, and angiotensin II, induce proliferations in these cells, correlating *Egr-1* mRNA and protein induction with cell proliferation.

Strong evidence for a role for *Egr-1* in proliferation also comes from studies with mouse skeletal muscle Sol8 cells (*20*). Although *Egr-1* message was induced in response to mitogenic stimuli (such as basic fibroblast growth factor, PDGF BB, and fetal calf serum), differentiative stimuli (insulin), and other agents that caused neither proliferation nor differentiation, Egr-1 protein could be detected only in response to mitogenic cues. Translation of *Egr-1* may be uncoupled from transcriptional induction, as was in fact suggested by earlier studies with human fibroblasts (*21*). Although interferons α and γ, tumor necrosis factors α and β, and epidermal growth factor induced *Egr-1* message levels to a similar extent, the amount of *Egr-1* translated varied with the mitogenicity of the inducing agent. Cao *et al.* (*21*) suggest that the mechanism of translational regulation may be through the phosphorylation of cap-binding protein (eIF-4E). Phosphorylation of this factor, which promotes cellular protein synthesis, is enhanced by the mitogenic agents EGF and tumor necrosis factor (TNF) but not interferon (IFN) (*21*). Together these studies present an intriguing correlation between the translatability of *Egr-1* message and the strength of the mitogenic inducing signal. Given the translational block in Egr-1 production induced by insulin in Sol8 cells, any role for Egr-1 in differentiated muscle must assume a function for the abundantly expressed *Egr-1* message, perhaps within its 3′ UTR (*22*). In light of these results, the assumption that *Egr-1* mRNA levels correlate with protein levels implicit in many studies of *Egr-1* induction must be reexamined.

Finally, recent work (*23*) suggests a role for Egr-1 in the regulation of astrocyte growth. Endothelin 3 (ET-3), a potent growth regulator in these cells, stimulates Egr-1 and basic fibroblast growth factor expression. An antisense oligonucleotide to *Egr-1* blocked ET-stimulated thymidine uptake and bFGF gene transcription. Moreover, an antisense oligomer to the bFGF gene significantly blocked ET-stimulated thymidine incorporation. These studies point to a causal role for *Egr-1* induction in the proliferation of astrocytes and suggest that the bFGF gene may be a relevant physiological target gene.

B. Induction during Development and Differentiation

In the adult mouse, high levels of *Egr-1* mRNA are seen in brain, thymus, heart, muscle, and lung. In particular, the high level of expression in the brain is located in the cerebral cortex and hippocampus (*14*). Lower levels are detected in kidney, spleen, and most other tissues, with very low levels in liver (*6, 13, 14*). A similar pattern of expression has been observed in the adult rat: *Egr-1* is most abundant in brain and adrenal gland, and is also highly expressed in superior cervical ganglia and lung (*24, 25*).

During development, a single *Egr-1* transcript is predominantly expressed in cortex, midbrain, and cerebellum; in bone, cartilage, and muscle; and at several sites of epithelial–mesenchymal interactions. Studies in the developing rat suggest a role for *Egr-1* in postnatal maturation of the brain: *Egr-1* levels are low in neonatal and early postnatal brain, but increase dramatically at later times and in the adult animal, with highest levels detected in the cortex (*25*). In the developing mouse, *Egr-1* expression in 14.5- and 17.5-day fetal skeleton parallels *c-fos* expression, suggesting a role for these coregulated genes in skeletal development. *Egr-1* expression is correlated with the onset of ossification (about day 14.5) and is localized to regions of the embryo undergoing substantial bone formation, including the membranous and alveolar bones of the head and the periosteal and endochondral ossification sites of the developing long bones (*26*). Like *c-fos*, *Egr-1* is expressed in cartilage at the articular surfaces of joints and in the interstitial

cells that lie in between these elements. In addition, high-level *Egr-1* expression is seen in developing striated muscle, showing a patchy distribution. Finally, it has been suggested that *Egr-1* may respond to signals that mediate epithelial–mesenchymal interactions during organogenesis: expression is localized to ectodermally derived cells of the inner root sheath in young whisker follicles, in the underlying mesenchymal component of developing salivary and nasal glands of the mouse, as well as the mesenchymal component of the developing tooth (26). This initial patterning during tooth organogenesis requires primary signals derived from the dental epithelium. Importantly, recent *in vitro* reconstitution experiments demonstrate that purified bone morphogenetic protein 4 (BMP-4) can substitute for dental epithelium in inducing morphogenetic changes in the mesenchyme and in up-regulating *Egr-1* expression (27). In summary, the developmental profile of *Egr-1* is consistent with a role for it in brain maturation, in skeletal development, and in response to epithelial–mesenchymal interactions.

In several cell types, a rise in *Egr-1* expression is correlated with differentiative processes, in particular in cardiac, neural, osteoblast, and monocytic differentiation. Differentiation of P19 embryonal carcinoma cells into cardiac muscle, or nerve and glial cells, is induced in the presence of dimethyl sulfoxide (DMSO) or retinoic acid, respectively. In response to either, a biphasic pattern of *Egr-1* expression is seen. A transitory increase after 3 days of treatment is followed by high sustained levels of *Egr-1* expression after 14 days in culture (6). The expression of *Egr-1* in adult heart and brain is consistent with its prolonged response, pointing to a role for it in these differentiated cell types (6, 25). Neuronal differentiation can also be modeled on the rat pheochromocytoma cell line PC12. Nerve growth factor (NGF) causes an initial mitogenic response in PC12 cells, followed by growth arrest and differentiation into sympathetic neuronlike cells with extended neurites. *Egr-1* responds rapidly to NGF in PC12 cells, as to other growth factors; however, the expression is not transient, remaining high for up to 6 days (6, 8).

Finally, retinoic acid induces the differentiation of rat calvarial preosteoblastic RCT-1 cells. *Egr-1* is induced rapidly and transiently by retinoic acid in RCT-1 cells or primary cultures of embryonal calvarial cells, but not in the most mature RCT-3 line, which already expresses many osteoblastic markers (29). These observations, together with the expression of *Egr-1* in developing bone and cartilage described above, support a role for Egr-1 in osteoblast differentiation (26, 29).

As described above, *Egr-1* induction has been correlated with the onset of differentiation in several cell types. In particular, monocytic differentiation of U-937 and HL-60 myeloid leukemia cells induces *Egr-1* expression (30, 31). Interestingly, dexamethasone, an inhibitor of monocytic differentiation, blocks the *Egr-1* induction (31). Recent exciting results with myeloid cells provide the first demonstration that *Egr-1* expression is *necessary* for differentiation (32). The human myeloid leukemia cell line HL-60 can be induced to differentiate along either macrophage or granulocyte lineages by treatment with phorbol or DMSO, respectively. *Egr-1* expression is seen exclusively on induction of macrophage differentiation in HL-60 cells and primary myeloblasts. *Egr-1* antisense oligomers added to the culture medium prevent macrophage differentiation, and constitutive expression of Egr-1 limits the differentiative capacity of HL-60 cells such that these multipotent cells can be longer be induced for granulocyte differentiation (32). These results convincingly demonstrate that Egr-1 expression is essential for and restricts differentiation along the macrophage lineage. Mapping of the human *EGR1*[3] gene to chromosomal locus 5q31.1 is particularly intriguing with respect to these studies. The human EGR1 locus has been localized to a 2.8-megabase region defined by overlapping chromosomal deletions from patients with therapy-related acute myeloid leukemia (6, 33). The suggestion that *Egr-1* is a myeloid tumor-suppressor gene is consistent with a role for *Egr-1* in promoting myelogenesis.

C. Induction by Tissue or Radiation Injury

In a third context, *Egr-1* is induced in response to tissue or radiation injury. Ischemic injury to the kidney results in alterations in epithelial cell polarity, tissue damage, and cell death. Restoration of differentiated function after ischemic injury sets the kidney apart from the heart and brain, two organs that are irreversibly damaged by oxygen deprivation. Ischemic injury to rat kidney followed by reoxygenation induces a transient 30-fold increase in *Egr-1* expression that does not require protein synthesis. Moreover, because induction requires reoxygenation, *Egr-1* is not induced by the injury *per se*, but may rather act in response to postischemic events to mediate the subsequent processes of cell differentiation or proliferation (34). A second example of *Egr-1* induction as a consequence of cell injury is the cellular response to X-ray irradiation. Ionizing radiation has pleiotropic effects, including growth arrest, the repair of damaged DNA, and proliferation. *Egr-1* responds by a transient induction within 0.5 to 3 hours of exposure to X-rays in the absence of protein synthesis (35).

D. Induction in Neuronal Signaling

Immediate-early genes, by analogy to their part in the mitogenic response, may also play an important role in stimulus-transcription coupling in neurons (36). Several lines of experimentation indicate that immediate-early

[3] EGR is the human factor; Egr is the mouse or rat factor.

Response	Stimulus	Cell type	References
Mitogenic	serum	fibroblasts	3, 6, 9, 13
	PDGF, EGF, FGF	fibroblasts	3, 6, 9, 13, 42
	insulin	hepatocytes	3
	phytohemagglutinin	peripheral blood lymphocytes	5, 22
	anti-μ	B lymphocytes	17
	adenosine diphosphate	kidney epithelial cells	9
	PDGF, vasopressin	kidney mesangial cells	19, 114
	serum	skeletal muscle Sol8 cells	20
	bFGF, PDGF BB	skeletal muscle Sol8 cells	20
	partial hepatectomy	liver	9, 179
	GM-CSF, LPS	peritoneal macrophages	132, 133
	endothelin	astrocytes	23
	angiotensin II	vascular smooth muscle cells	126
Hypertrophic	endothelin, angiotensin II	myocyte	103, 117
Differentiative	NGF	pheochromocytoma PC12 (neural)	6, 8
	retinoic acid, DMSO	embryonal carcinoma P19	6
	retinoic acid	embryonal cerebral cells, NCT-2 (neuroblast)	29
	TPA, DMSO	myeloid leukemia HL60 and U-937	30, 31, 32
Tissue radiation injury	ischemia	kidney	34
	ionizing radiation	293, BQ-293	25, 53, 116
Neuronal excitation	potassium ions	PC12 (depolarization)	36, 6, 37
	metrazole	seizures in vivo	6
	NMDA	hippocampus	40
	visual stimuli	visual cortex	38, 39
	electroconvulsive shock therapy, dopamine receptor activation, opiate withdrawal	CNS	42
	peripheral nervous system	sciatic nerve transection	42
Other	ras	MDCK, LLC-PK, renal epithelial cells	118

FIG. 1. Biological processes in which *Egr-1* expression has been described.

genes, including *Egr-1*, participate in the rapid response of neurons to transsynaptic stimuli. *In vivo*, *Egr-1* levels increase rapidly in the brain following seizure activity, with kinetics similar to c-*fos* (6). Membrane depolarization of PC12 cells by treatment with potassium chloride also results in rapid and transient induction of *Egr-1* (6, 37). In dark-reared cats, a brief 1-hour visual stimulation causes dramatic and transient induction of *Egr-1*, c-*fos*, and *junB* mRNAs that are specific to the visual cortex, i.e., absent from the frontal cortex. The magnitude of the induction, greatest in young animals, is consistent with the idea that *Egr-1* expression plays a fundamental role during the critical period of development in the visual cortex (38, 39). A role for *Egr-1* in postnatal maturation of the brain is supported by the dramatic increase in *Egr-1* message levels in all sections of postnatally developing rat brain,

especially cortex (25). Finally, high-frequency stimulation of the perforant path–granule cell synapse results in induction of *Egr-1* in the postsynaptic cells. The response of *Egr-1* is highly reproducible, as compared to the variable response of other immediate-early genes. Interestingly, induction of *Egr-1* is correlated with long-term potentiation, because both responses require the *N*-methyl-D-aspartate receptor and a stimulus of similar frequency and intensity (40). Additional studies show *Egr-1* induction following electroconvulsive shock therapy, D1 dopamine receptor activation, and opiate withdrawal (41). Transient *Egr-1* induction has also been noted in the peripheral nervous system, e.g., sciatic nerve transection provokes Egr-1 protein increase in neurons of the spinal dorsal horn (42). These studies, and the expression of *Egr-1* in developing and adult brain and in the peripheral nervous system are consistent with a role for Egr-1 in neurophysiological processes.

This summary of the contexts in which *Egr-1* is expressed emphasizes the diversity of signals that induce *Egr-1* (Fig. 1). *Egr-1* is induced by mitogenic stimuli in all cell types; during differentiation of nerve, cardiac, bone, and myeloid cells; after tissue injury due to ischemia or irradiation; and by signals that result in neuronal excitation, such as membrane depolarization or brain seizures. There has been one demonstration, in the differentiation-inducible HL-60 cell line, of a phenotype resulting from inappropriate Egr-1 expression (32). In addition to promoting and restricting differentiation of myeloid precursors along the macrophage lineage, the enormous complexity of the Egr-1 response hints that this protein may play diverse roles in different cellular contexts.

IV. Proximal Events

A. Second Messengers

Two strategies have yielded insight into the complex regulation of the *Egr-1* gene: activation or inhibition of specific second-messenger pathways and a molecular genetic dissection of the *Egr-1* promoter. Multiple intracellular pathways appear to contribute to the regulation of *Egr-1* expression. Both protein-kinase-C (PKC)-dependent and -independent mechanisms are integral in linking extracellular signals to transcriptional activation of *Egr-1*. Clearly, the PKC pathway can relay extracellular stimuli to a nuclear response resulting in *Egr-1* induction, because direct activation of the pathway by phorbol ester (TPA) induces *Egr-1* (5, 43). In addition, non-PKC pathways also play a role: fibroblasts rendered deficient in PKC signaling by long-term exposure to phorbol retain a robust *Egr-1* response to serum and epidermal growth factor (47).

In the response of *Egr-1* to tumor necrosis factor and interferon in human fibroblasts, the PKC pathway appears instrumental. Treatment with H7 (a nonspecific inhibitor of protein kinases including PKC) or the PKC inhibitor staurosporine effectively blocks much of the *Egr-1* response. The selective inhibitor of cyclic-nucleotide-dependent protein kinases, HA1004, does not modify the *Egr-1* response (*21*). Stimulation of B lymphocytes with phorbol or the PKC agonist SC-9 also up-regulate *Egr-1* expression, implying that surface immunoglobulin (Ig)-generated signals work through PKC. Evidence for the PKC pathway as a requisite component of anti-μ induction of *Egr-1* comes from studies with inhibitors of PKC. A prior treatment with either H7 or sangivamycin, effective inhibitors of PKC, blocks the increase in *Egr-1* mRNA levels in response to anti-μ. Again, the cyclic-nucleotide-dependent protein kinase inhibitor HA1004 had no effect. These studies demonstrate that activation of PKC is involved in coupling surface Ig stimulation in B lymphocytes to the transcriptional response of the *Egr-1* gene (*44*).

The PKC pathway appears fundamental in mediating *Egr-1* induction in response to X-irradiation. First, prolonged stimulation with micromolar concentrations of phorbol depletes PKC and virtually blocks the X-ray inducibility of Egr-1 in SQ20B cells. In addition, pretreatment with the inhibitor H7 but not HA1004 markedly attenuates the X-ray inducibility of *Egr-1* in SQ20B or 293 cells (*35*).

In contrast, an intracellular pathway involving c-Raf plays a central role in the v-Src induction of *Egr-1*. c-Raf-1 is a serine-threonine protein kinase, and v-Raf up-regulates the *Egr-1* promoter. Moreover, expression of a kinase-defective mutant of c-Raf-1 blocks induction of *Egr-1* upon regulation of the *Egr-1* gene.

B. Egr-1 Promoter Analysis

The architecture of the *Egr-1* promoter has been described by several groups who have cloned the murine (*14*, *46*), rat (*47*), and human *Egr-1* genes. In particular, the coregulation of c-*fos* and *Egr-1* in several contexts has prompted a comparison of their promoter sequences. Six CC(W)6GG elements (CA_nG boxes), the functional core of the serum response element (SRE), are present in the *Egr-1* promoter; however, none of these potential SREs shares the extended symmetry outside of the core sequence that typifies the c-*fos* SRE (*48*). In addition to the CA_nG boxes, putative regulatory elements in the *Egr-1* promoter include cAMP response elements, AP1, CREB, and Sp1 sites as well as a CCAAT box and TATA motif (*14*, *46*, *47*, *49*) as illustrated in Fig. 2.

The demonstration that 1 kb of murine 5′ sequence confers serum and phorbol responsiveness to a CAT reporter in mouse fibroblasts opened the door to delineation of the functional elements (*14*, *50*, *51*). Similarly, NGF inducibility was observed with the sequence from −532 to +100 of the rat gene in PC12 cells (*47*). Deletion analysis of the *Egr-1* promoter showed that a construct with sequence to −594 (and all six CA_nG boxes) retains full serum inducibility whereas deletion to −166 (with the two proximal CA_nG elements) has partial serum responsiveness as compared to a minimal promoter construct. Moreover, synthetic constructs with a single *Egr-1* CA_nG box confer serum inducibility on the heterologous thymidine kinase promoter (*49*). These results show clearly that the decanucleotide inner core of the previously defined c-*fos* SRE functions as a serum response element in the *Egr-1* promoter. In a gel-shift assay, the core *Egr-1* SRE can compete for binding against the c-*fos* SRE with its more extensive dyad symmetry. And like the c-*fos* SRE, the *Egr-1* CA_nG boxes bind to in-*vitro*-translated serum response factor.

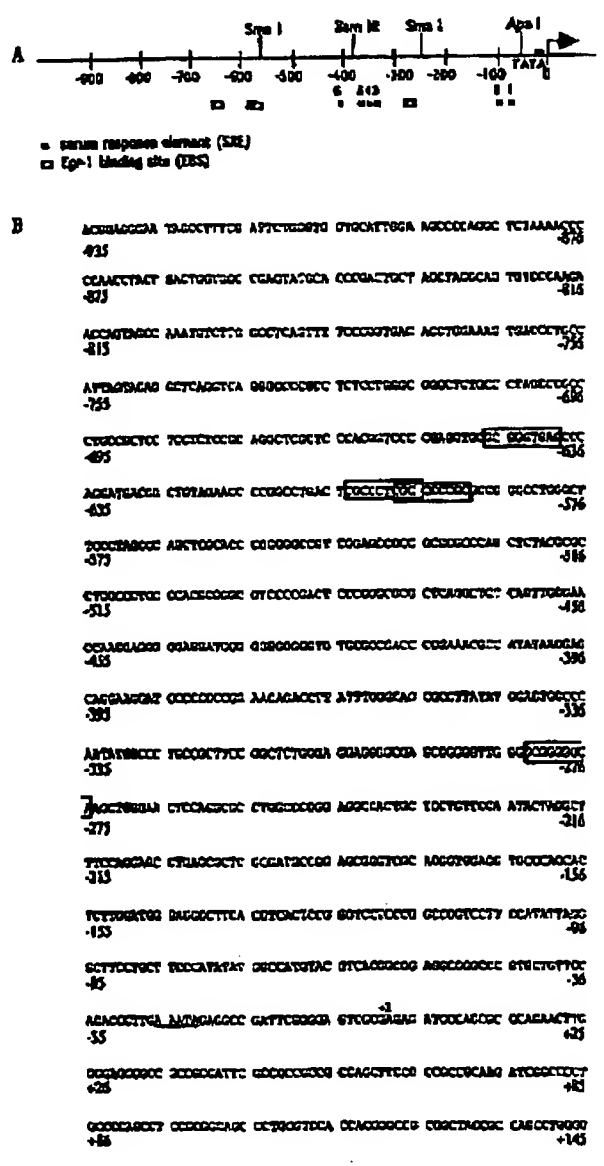
Further experiments with synthetic constructs indicate that tandem copies of the CA_nG boxes are more strikingly inducible than an individual element (*49*). Given these observations, the greater serum inducibility of *Egr-1* versus c-*fos* may be explained by the multiple elements in the *Egr-1* promoter as compared to the single SRE regulating c-*fos* expression. The CA_nG box appears to play a central role in the broad responsiveness of *Egr-1* to mitogens, because this motif directs induction by PDGF, phorbol, v-*src*, and v-*fos*, as well as serum (*16*, *49*, *50*, *52*). These elements, especially the three most 5′ ones, are also responsible for the activation of *Egr-1* by ionizing radiation (*53*).

Finally, the CA_nG boxes in the *Egr-1* mediate the down-regulation of *Egr-1* transcription following mitogenic stimulation. In particular, the Fos protein effects this transcriptional repression; Fos mutants lacking a leucine zipper function as well in this assay, and the C-terminal region of Fos is sufficient for this function (*51*).

V. Distal Events

A. Characterization of the Egr-1 Protein Product

Immediate-early genes encode several types of proteins, including growth factors, growth factor receptors, cytoskeletal proteins, and transcription factors. Sequence analysis of Egr-1 revealed a protein with three tandemly repeated Cys_2His_2 zinc-finger motifs that presage the function of this protein (*5*, *8*, *13*, *14*). The zinc finger (see below), a highly conserved eukaryotic DNA-binding motif, is a compact domain that uses conserved pairs of cysteine and histidine residues to coordinate a central zinc ion (*54*). The importance of the *Egr-1* gene product is also suggested by the conservation

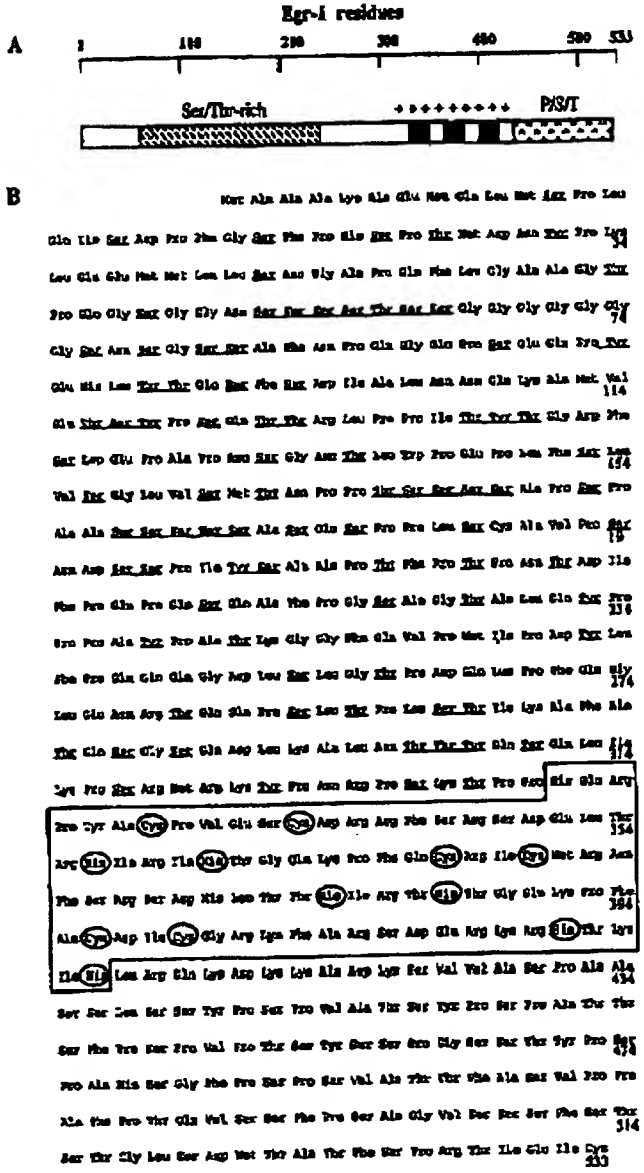


of the coding sequence across vertebrate evolution: human (7), rat (8), mouse (6, 13, 14), chicken (11), and zebrafish (55, 56) cDNAs are highly homologous.

From the deduced amino-acid sequence of Egr-1 protein, several interesting features have been predicted (Fig. 3A). Basic residues cluster in the three zinc fingers and adjacent sequence. The amino-terminal 300 amino acids are rich in proline (14%) and serine/threonine residues (34%). Several stretches of five to seven consecutive serine or threonine residues are present with one series of seven serine/threonine residues followed by seven glycines (Fig. 3B). It has been noted that the repeating trinucleotide motifs that encode these poly(aminoacid) stretches are similar to those whose expansion has been implicated in human disease (55, 57, 58). The region on the carboxy-terminal side of the zinc-finger motifs is also rich in proline (15%) and serine plus threonine (37%), but this region is distinguished by a repeated motif of eight amino acids with the consensus Ser/Thr-Ser/Thr-Phe/Tyr-Pro-Ser-Pro-X-X. The composition of this reiterated sequence is reminiscent of the heptapeptide repeat in the carboxy-terminal domain of the RNA polymerase II large subunit (59). The proline-rich regions of Egr-1 are predicted to lack α-helical secondary structure, whereas the high content of serine, threonine, and tyrosine residues suggests that Egr-1 may be phosphorylated.

Characterization of the *Egr-1* gene product showed it to encode a short-lived protein with an anomalous electrophoretic mobility of 80–82 kDa. In fibroblasts, Egr-1 protein is rapidly induced by serum, accumulating within 30 minutes and reaching maximum levels at 1–2 hours poststimulation (50). Consistent with its putative DNA-binding function, immunocytochemistry and cell fractionation studies show that Egr-1 is located in the nucleus (50, 60, 61). Studies (60) have characterized the rat homolog in PC12 cells with several antisera directed against various regions of the protein. In particular, a truncated species of 54 kDa is cytoplasmic. This 54-kDa species is recognized by antisera directed against the basic region immediately 5′ of the first zinc finger but not by sera against a C-terminal peptide. These results were an early indication that sequences within or C-terminal to the zinc-finger

FIG. 2. The 5′ upstream sequence of the murine *Egr-1* gene. (A) Schematic of promoter depicting putative regulatory elements. The positions of the six serum response elements within approximately 1 kb of promoter sequence are depicted as darkened boxes. The locations of the Egr-1 binding sites within the promoter are indicated as open boxes. (B) Nucleotide sequence of the *Egr-1* promoter. [Reprinted from NARes (Ref. 46) by permission of Oxford University Press.] The nucleotides are numbered from the cap site, which is +1. The putative TATA element is underlined and three Egr-1 binding sites in the 3′ promoter region are boxed.



domain may participate in nuclear targeting (60). The *Egr-1* gene product is also phosphorylated: alkaline phosphatase converts the two closely spaced Egr-1 species seen on SDS–PAGE analysis of NGF-stimulated PC12 cells to the faster migrating form (50, 60). Immunoprecipitation of Egr-1 from phosphate-labeled HeLa cells and subsequent analysis of phosphoaminoacid content indicate that the phosphorylation is on serine (62).

B. DNA-binding Activity of Egr-1

Hundreds of eukaryotic transcription factors share the highly conserved DNA-binding motif known as the zinc finger. First identified as a compact zinc-binding domain in the *Xenopus* transcription factor IIIA (54), this well-conserved motif also occurs in the yeast proteins SWI5 and ADR1, *Drosophila* factor Krüppel and Hunchback, and mammalian regulatory proteins such as the testis-determining factor ZFY, the enhancer-binding protein Sp1, and the Wilms tumor suppressor WT1. TFIIIA-like fingers are distinguished by pairs of conserved cysteine and histidine residues and are evolutionarily and structurally distinct from the cysteine-rich zinc-binding motifs in the steroid receptors and in the yeast factor GAL4. A variable number of tandem repeats of this domain of 28–30 amino acids act in concert to recognize a specific DNA sequence (63).

Residues fundamental to the structural integrity of the finger domain are conserved among all Cys_2His_2 zinc-finger proteins whereas other amino acids involved in base sequence discrimination may be unique or confined to a subset of this large family of proteins. Pairs of cysteine and histidine residues are absolutely conserved as are usually the hydrophobic amino acids phenylalanine and leucine (Fig. 4A). The region connecting the histidine of one finger to the cysteine of the following finger, designated the H–C link, has the highly conserved consensus His-Thr-Gly-Glu-Lys/Arg-Pro-Phe-Tyr-X-Cys (67). In addition, three variable residues, discussed below, appear to participate in sequence-specific interactions with DNA.

NMR and crystallographic studies suggest that each zinc-finger motif consists of an antiparallel β-sheet that includes the two consensus cysteines, and an α-helix that contains the two conserved histidine residues (Fig. 4B). Each

FIG. 3. Schematic structure and amino-acid sequence of the Egr-1 protein. (A) Structural features of Egr-1. Each zinc-finger motif is designated by a black bar. The basic region of Egr-1 is indicated (+++). The serine/threonine-rich N-terminal domain of Egr-1 is shown on the left and the proline/serine/threonine-rich C-terminus (P/S/T) is on the right. (B) Coding sequence of murine Egr-1. [Reprinted with permission from Ref. 6, copyright 1986 Cell Press.] The three zinc-finger motifs are enclosed. Conserved cysteine and histidine residues in the zinc fingers are circled. Serine, threonine, and tyrosine residues in the N-terminal domain are underlined.

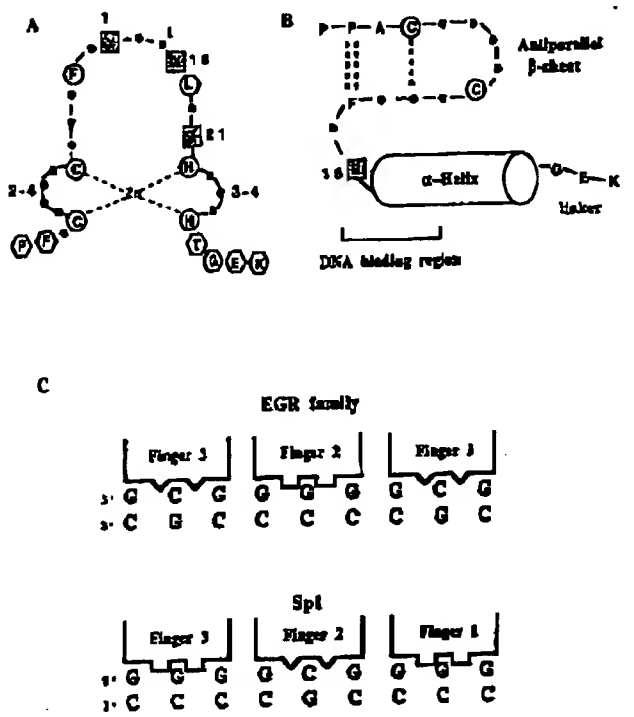


FIG. 4. The zinc finger as a modular DNA-binding motif. (A) Zinc-finger consensus residues. Invariant cysteine (C) and histidine (H) residues that coordinate a zinc ion are circled, as are the conserved hydrophobic residues phenylalanine and leucine. Residues that are part of the highly conserved His–Cys link are enclosed by hexagons. Amino acids that determine the sequence-specificity of binding are shown in the shaded boxes. (B) Diagram of zinc-finger folding (from *163*). Each zinc-finger domain is composed of a β-sheet and an α-helix. Hydrogen bonds are depicted as dotted lines. (C) Each zinc finger contacts a three-nucleotide subsite. [Reprinted with permission from *Nature* (Ref. 66), copyright 1991 Macmillan Magazines Limited.] Fingers 1 and 3 of Egr-2 are postulated to bind the same three-nucleotide subsite as finger 2 of Sp1.

zinc finger incorporates these secondary structures into a compact globular domain with the invariant cysteine and histidine residues coordinating a central zinc ion. A hydrophobic core including the conserved phenylalanine and leucine residues and the first histidine stabilizes the domain. In a manner similar to prokaryotic helix–turn–helix motifs and eukaryotic homeodomains, the α-helix of the zinc finger lies within the major groove of DNA. Multiple interactions between amino-acid side-chains of the helix and DNA base-pairs combine to discriminate among nucleic-acid sequences (*64*).

In a search for the DNA element recognized by Egr-1 among fragments derived from the 5′ upstream flanking sequence of the *Egr-1* gene (*49*), it was found that autoregulation by other immediate early genes such as *c-fos* and cycloheximide superinduction of *Egr-1* was consistent with the hypothesis that *Egr-1* regulates its own expression. Using gel mobility shift assays with Egr-1 protein purified from bacteria, specific binding to one promoter fragment was observed (*49*). DNase-I footprinting identified the sites of contact, revealing that Egr-1 binds the 9-bp sequence GCG-GGG-GCG. Further gel shifts comparing the affinity of this sequence to sites altered at various positions generated the consensus sequence: GCG-KGG-GCG (*49*). Gel shift assays with zinc-chelating agents were utilized to demonstrate the requirement for zinc cations to effect DNA binding (*50*).

Two types of experiments support a similar model for the determination of DNA-binding specificity by EGR fingers and proteins with related zinc-finger domains (reviewed in *65*). Mutagenesis experiments were guided by the similar but distinct zinc-finger domains of Sp1 and Egr-2, a gene whose three zinc fingers are identical to those of Egr-1 except for four conservative amino-acid substitutions. These mutagenesis studies foreshadowed the results obtained by cocrystallization of the Egr-1 zinc-finger domain and its cognate binding site. It has been observed that Sp1 and Egr-2 each contain three zinc fingers and bind to a (G+C)-rich 9-bp binding site (*66*). If each motif interacts with DNA in an analogous manner, then a zinc finger is predicted to contact 3 bp of DNA. Furthermore, comparison of the Egr consensus GCG-GGG-GCG to the Sp1 consensus GGG-GCG-GGG suggested that fingers 1 and 3 of Egr-2 might have the same specificity-determining residues as finger 2 of Sp1 (Fig. 4C). Fingers 1 and 3 of Egr-2 share Glu^{18} and Arg^{21} with finger 2 of Sp1 (Fig. 5A). In addition, finger 2 of Egr-2 and fingers 1 and 3 of Sp1 each have a histidine residue at position 18 of the finger. It was predicted (*66*) that the residues at positions 18 and 21 discriminate between GCG or GGG subsites. In accordance with this hypothesis, mutagenesis of Egr-2 finger 2 residues His^{18} to Glu and Thr^{21} to Arg created a protein that did not bind the *Egr-2* cognate sequence but instead recognized the novel sequence GCG-GCG-GCG (*66*).

The mode thus constructed (*66*), in which variable residues at positions 18 and 21 were postulated to be the determinants of base-sequence discrimination, has been substantiated (*64*). Solution of the Egr-1 zinc-finger domain–DNA crystal structure provided a framework for understanding how proteins with tandemly repeated Cys_2His_2 zinc fingers interact with DNA.

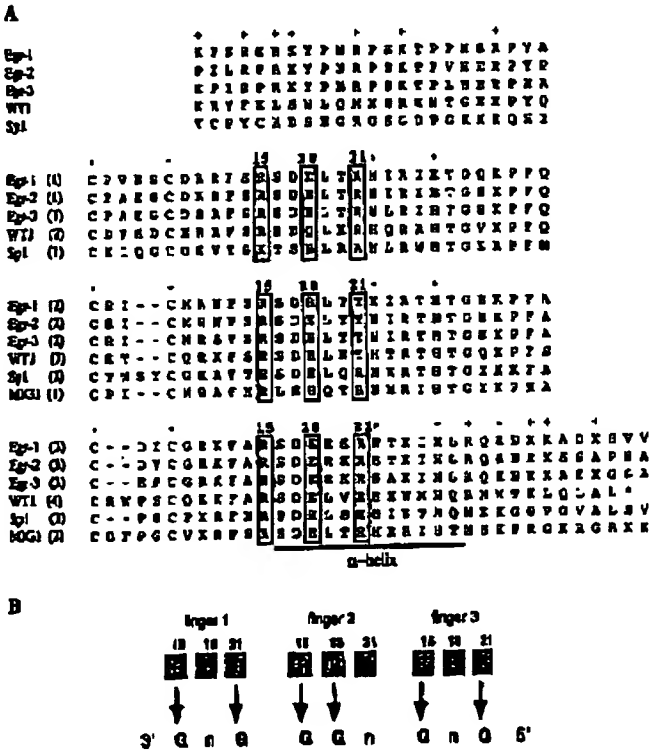


FIG. 5. DNA-binding domains of the EGR family. (A) Comparison of DNA-binding domains related to Egr-1. Zinc fingers and the flanking sequence of murine Egr-1, human Egr-2, human Egr-3, the Wilms tumor gene WT1, Spl, and the yeast protein MIG1 are aligned for comparison. The position of each finger motif is noted in parentheses. Conserved cysteine and histidine residues are marked (·). The helical region is underlined and the residues important in determining binding specificity are enclosed. Conserved basic residues flanking the zinc-finger domains are denoted (+). (B) Residues determining sequence specificity of Egr-1 binding. [Adapted with permission from R. E. Klevit, Science 253, 1367 (1991) (Ref. 65). Copyright 1991 American Association for the Advancement of Science.] The Egr-1 zinc-finger domain interacts with the guanine-rich strand of DNA in an antiparallel manner. Fingers 1 and 3 contact the same 3-bp subsite. Arrows represent specific interactions between arginine or histidine residues and the guanine bases.

This structure showed that each finger has a similar relation to the DNA and interacts primarily with a 3-bp subsite. The α-helix of each finger fits directly into the major groove so that residues in the amino-terminal part of the helix

are available for hydrogen-bonding interactions with the base-pairs. The β-sheet is on the backside of the helix away from the base-pairs; the second β-strand of the sheet contacts the sugar-phosphate backbone, serving to orient the α-helix in relation to the DNA. An arginine immediately preceding the helix makes important DNA contacts as do the second, third, and sixth residues of the α-helices. Each of the hydrogen bonds are made to guanines of the G-rich strand of the DNA. The orientation of the three fingers is antiparallel with respect to the G-rich strand; that is, the 5′-most subsite of the sequence is recognized by the carboxy-terminal finger. In addition, the α-helix lies in the major groove in an antiparallel manner, so that the carboxy-terminal portion of each helix interacts with the 5′-most base of each subsite (64).

A select number of residues at defined positions in each finger interact specifically with the DNA. The residue preceding the α-helix and the third and sixth residues of the helix make the specific contacts. In each finger of Egr-1, Arg^{18} precedes the helix and Asp^{17} is the second residue of the helix (Fig. 5A). The arginine hydrogen bonds through its long side-chain with a guanine at the third position of each subsite and is stabilized by the conserved aspartate residue. Thus, the third position, G, is common to each subsite and is recognized in an identical manner (Fig. 5B). The third residue of the helix varies between fingers; Glu^{20} is present in fingers 1 and 3 whereas a histidine is present at the same position of finger 2. The structure solved by Pavletich and Pabo (64) shows that these glutamate residues do not contact the DNA. In contrast, the histidine of finger 2 participates in a hydrogen bond with the guanine in the center of its subsite. The sixth residue of the helix, Arg^{21} in fingers 1 and 3, forms a specific bond with the guanine occupying the first position of the subsite. A threonine, which is the sixth residue in the helix of finger 2, is incapable of this interaction (64).

In summary, a relatively simple pattern has emerged for Egr-1:DNA recognition from this work (64). The common arginine immediately preceding the helix specifies a guanine at the third position of each subsite. The third residue of the helix may contact the middle base of the subsite, and the sixth residue of the helix may contact the first base of the subsite. Moreover, the Egr-1 zinc fingers utilize only arginine or histidine residues to contact guanines; in the absence of these amino acids, such as Glu^{18} in fingers 1 and 3 and Thr^{21} in finger 2, there is no specific recognition of the DNA sequence (64).

A complementary analysis (49) of variant *Egr-1* binding sites has confirmed the lack of specific interactions with the fourth nucleotide: in gel shift assays, GCG-TGG-GCG competed as well as GCG-GCG-GCG. However, the sequences GAC-GCG-GCG and CCG-GGC-GAC were not efficient competitors, showing that not all nucleotides are permissible at positions 2

and 8 (49). Although no specific contacts were observed at these positions in the Egr-1:DNA cocrystal (64), it is possible that substitution of the bulkier adenosine for cytosine is disruptive at these positions.

These results (64, 66) emphasize the modularity of the zinc-finger motif in which each zinc-binding domain recognizes a three-nucleotide sequence. In particular, an implicit assumption has been that each finger makes an equal contribution to the overall affinity of binding. A complementary in vivo mutational analysis of the Egr-1 zinc-finger domain hints that each finger may not make the same contribution to binding. Specifically, many more DNA-binding impaired mutants with alterations in the second finger rather than in the first or third can be recovered (67). Moreover, the two His-Cys links connecting the finger motifs also showed a disparity in the number of DNA-binding mutants recovered. The second linker was mutated 17 times whereas the first was altered three times, suggesting that the linkers may not play identical roles in orienting the fingers (67).

The recognition code outlined by the crystallographic studies (66) indicates similar interactions for all three fingers of Egr-1 and implies that other Cys_2 His_2 zinc-finger proteins will use residues at analogous positions to make their base-specific contacts. Studies with the *Drosophila* finger protein Tramtrack reveal an extension to the formula derived from Egr-1 DNA–protein interactions whereby residues at three positions determine DNA binding specificity. The first finger of Tramtrack uses an additional amino-acid contact to recognize its DNA binding site (68). In conclusion, the model developed from Egr-1 studies will generalize to some other zinc-finger proteins, but it does not describe the complete repertoire of all possible protein–DNA contacts in Cys_2 His_2 zinc-finger proteins.

C. Structure–Function Analysis

1. Defining the Egr-1 *Trans*-activation Domains

Definition of a DNA-binding site for Egr-1 set the stage for assessing whether Egr-1 could regulate transcription through the GCG-GGG-GCG sequence. Data from transient transfection assays shows that Egr-1 can activate a minimal promoter with multiple *Egr-1* binding sites 10-fold in a dose-dependent manner (62, 69). Like classical transcription factors, the organization of Egr-1 is modular in nature, with functional domains that are structurally independent and able to confer activity on heterologous proteins. We and others have used deletion analysis and gene fusions to dissect the functional domains of Egr-1, delineating modular activation, repression, and nuclear localization activities.

Deletion analysis of murine Egr-1 indicates that the extensive serine- and threonine-rich N-terminal domain is a robust transcriptional activator. A

constructive approach shows that several Egr-1 activation sequences are independent domains capable of functioning in a heterologous context when fused to the DNA-binding domain of the yeast factor GAL4. Residues 3–281, or subregions from 3 to 138 or 138 to 281, activate transcription 100-fold as GAL4 fusions (70). Deletion analysis of the rat homolog of Egr-1 further suggests that residues 13–38 and 223–264 may be most important for the activation function (57). The N-terminal domain is 30% serine/threonine/tyrosine rich over a span of 180 residues; the large size of the activation domain may contribute to its potency relative to the smaller, previously described serine/threonine-rich activator Pit-1/GHF-1 (71). Moreover, the *trans*-activation domain is impervious to mutation in that substantial deletions in the extensive N-terminal domain do not destroy transcriptional activity. Finally, work from several laboratories maps a weak trans-activation function to the C-terminus of Egr-1, which contains the octapeptide repeats reminiscent of the phosphorylated YSPTSPS reiterations in the carboxy-terminal domain of RNA polymerase II (57, 59, 70).

2. Localization of an Egr-1 Repression Domain

An unexpected result of deletion analysis is that a small internal deletion immediately 5′ of the zinc-finger domain (Δ284–330) enhances trans-activation some fivefold in HeLa cells. Western-blotting and gel-shift analyses showed that this superactivation cannot be explained simply by overexpression or enhanced DNA binding of the deletion derivative relative to full-length Egr-1. The superactivation observed with Δ284–330 is consistent with the loss of a region important for repression or for negatively regulating the trans-activation function of Egr-1. Further experiments have shown that *Egr-1* encodes a portable repression domain. Initial work demonstrated that a domain of 34 amino acids (281–314) can repress transcription 7- to 10-fold when fused to the GAL4 DNA-binding domain and assayed for effect on a reporter with five GAL4 binding sites. Repression by this compact domain was dependent on a DNA binding anchor (70). A further definition of the essential region showed that residues 281–304 repress and that residues 290–314 are inactive (72). This domain, highly conserved throughout vertebrate evolution (55), represents a novel motif distinct from the previously described alanine- and glycine-rich repression module in Krüppel (73, 74); the hydrophobic and proline-rich Even-skipped repressor (75); the glutamine-, alanine-rich factor Dr1; and the proline-, glycine-rich repressor of WT1 (76). In the Egr-1 repression domain, depicted in Fig. 5A, 7 of 24 residues are serine or threonine. In light of the fact that Egr-1 is known to be phosphorylated (14, 50, 60, 61), this raises the question of whether the Egr-1 repression function may be regulated by this modification (see below).

Repression by Egr-1 may involve an interaction with a cellular factor. A

competition assay showed that overexpression of Egr-1 amino acids from 266 to 301 results in a dramatic increase in activation from an Egr-1 molecule whose DNA-binding domain has been replaced with that of GAL4 (57). These results suggest that the region 266–301 is sufficient for an interaction with a titratable cellular factor that normally inhibits Egr-1 activity. A single isoleucine-to-phenylalanine substitution at position 290 renders the 266–301 domain nonfunctional. As predicted, this $Ile^{290}Phe$ mutation in the context of the native Egr-1 protein results in dramatic superactivation such that this variant activates about 15 times better than wild-type Egr-1. It is suggested that the cellular factor that interacts through this domain is present in a wide variety of mammalian cells, although apparently not in *Drosophila* Schneider cells because there is no superactivation in this cell type (57).

Elucidation of the mechanism of *Egr-1* repression has begun with the definition of the minimal promoter elements required. Initial work had demonstrated repression with an *Egr-1/GAL4* chimera on a reporter containing a portion of the thymidine kinase promoter with multiple protein-binding elements in addition to a TATA box. However, both *in vivo* and with an *in vitro* transcription assay using bacterially expressed fusion proteins, minimal promoter constructs containing only a TATA or initiator element in addition to binding sites to direct the *Egr-1/GAL4* chimera are sufficient for repression (72). Although these observations suggest that *Egr-1* repression is mediated by some type of interaction with the basal transcription machinery, preliminary experiments indicate that *Egr-1* does not directly bind to either TBP, TFIIB, or TFIIE *in vitro* (77). Therefore, the *Egr-1* repression domain may bind to one of the many other proteins involved in complex formation or to an associated protein, presumably the widely expressed cellular factor titrated by Rosso *et al.* (57).

The compact Egr-1 repressor is serine- and threonine-rich, and in particular Thr-269 has homology to known PKC phosphorylation sites (Fig. 5A). Phosphorylation is clearly not required for repression, because bacterially expressed Egr-1 efficiently represses transcription *in vitro* (76). This work is consistent with the suggestion that phosphorylation inactivates the *Egr-1* repression domain, preventing an interaction needed for the transcriptional inhibition. Importantly, an Ile-to-Phe mutation at the position analogous to Egr-1 residue 290 in the PKC substrate neurogranin makes it a better substrate for the kinase (78). The corresponding mutation in Egr-1, which may similarly promote phosphorylation on Thr-269, renders the repression domain nonfunctional (57). The role of phosphorylation may therefore be to enhance the ability of Egr-1 to work as an activator, by muting its repression function.

Egr-1 is one of only a small number of factors that contain modular domains capable of regulating transcription both positively and negatively. Other examples include the *Drosophila* factor Krüppel (74), YY1/NF-E1/δ (reviewed in 60), and the immediate-early factors Fos and Jun (81). This work provocatively suggests that native Egr-1 may be a bifunctional protein, capable of alternatively activating or repressing transcription. Such a property may be common to immediate-early genes to allow for versatility of effector functions. Posttranslational modifications as discussed above can be envisioned to enable complex factors such as these to regulate transcription either positively or negatively. In the case of Egr-1, we can speculate that Egr-1 may either activate or repress transcription, depending on whether it is induced in response to positive growth or to differentiation cues, or that Egr-1 may activate and repress multiple target genes depending on their promoter content, thereby mediating multiple transcriptional effects in response to a single inducing agent. HL-60 cell differentiation by phorbol may exemplify the latter type of bimodal Egr-1 function. Because Egr-1 expression both promotes macrophage differentiation and prevents granulocytic differentiation, the bifunctional role of Egr-1 may be to stimulate genes essential for macrophage differentiation while repressing genes required for specialized granulocytic functions.

3. Mapping the Egr-1 Nuclear Localization Signal

Consistent with its role as a transcriptional regulator, Egr-1 has been shown by several groups to be localized in the nucleus (50, 60, 67). Small molecules and proteins less than 40–60 kDa may passively diffuse across the nuclear pores into the nucleus, whereas larger proteins are targeted to the nucleus by an active, two-step process. The first step is a rapid, signal-dependent binding to the nuclear pore periphery, and the second step is a slower, ATP- and temperature-dependent translocation across the pore. In a number of nuclear proteins, the signal that specifies nuclear localization (NLS) is generally a short stretch of 6–10 amino acids characterized by basic residues as well as proline (reviewed in 82 and 83).

In Egr-1, basic residues cluster only in the three zinc fingers and adjacent sequences (Fig. 5), hinting that the karyophilic signal of Egr-1 resides here. Using subcellular fractionation/Western analysis or immunocytochemistry to analyze deletion derivatives of Egr-1, we have demonstrated that ΔN314 and ΔC430 are properly targeted to the nucleus, whereas ΔC314 is cytoplasmic. From these results, amino acids 315 to 429, encoding the three zinc fingers and adjacent basic sequences, appear essential for proper nuclear targeting. These results agree with early suggestions that the C-terminus of Egr-1 is required for nuclear localization (60).

A series of fusions of segments of Egr-1 to the large bacterial protein β-galactosidase were further used to show that the zinc-finger domain itself cannot function as an NLS. However, the zinc fingers in conjunction with

the 5′ basic sequence 315–330, but not the 3′ basic sequence, were sufficient to target the bacterial protein β-galactosidase to the nucleus. This 5′ basic stretch of residues 315–330, KPSRMRKYPNRPSKTP, is shared by other members of the EGR family, Egr-2 and Egr-3, which have conserved DNA-binding domains but generally diverge outside this region (Fig. 4A). Additional analyses showed that the entire zinc-finger domain is not required: either finger 2 or 3, yet not finger 1, could work with the 5′ basic sequence to form a bipartite NLS (70). Precedents for the incorporation of nuclear targeting signals within a DNA-binding domain include Fos (84); the progesterone receptor, in which the second finger but not the first functions as an NLS (85); GAL4 (83); and the homeodomain proteins α2 and Pit-1/GHF-1 (71, 86). Egr-1 may be a prototypical Cys_3His_1 zinc-finger protein whose DNA-binding and nuclear localization functions have coevolved as a composite domain rich in basic residues.

Other bipartite nuclear localization signals with two basic regions separated by a short variable spacer have been characterized in nucleoplasmin (87); SW15 (88); the *Xenopus* protein N1 (89); the steroid hormone receptors; and polymerase basic protein 1 of influenza virus (90). In addition, discontinuous nuclear targeting signals are found in adenovirus DNA-binding protein (91) and the yeast repressor α2, which has two nonhomologous signals, i.e., a basic NLS found at the N terminus as well as a signal located in the homeodomain (86, 92). In these proteins, as in Egr-1, the essential domains are discontinuous in the primary sequence, and it has been suggested that the two parts of the signal may mediate separate steps in nuclear accumulation (86). Several Egr-1–β-galactosidase mutants containing the 5′ basic sequence (but neither finger 2 nor 3 intact) and showing staining ringing the nucleus may contain the portion of the signal for binding to, but not translocation across, the nuclear pore (70).

Each of the assays used to define the Egr-1 NLS measured the equilibrium nuclear/cytoplasmic distribution of protein. Future kinetic analyses may reveal additional sequences required for prompt nuclear localization. Although a signal of seven predominantly basic amino acids suffices for the nuclear accumulation of SV40 T antigen over a period of hours, a more extensive sequence resulted in nuclear targeting within minutes (93). Serum-dependent nuclear import has been described for the immediate-early transcription factors c-Fos and, reportedly, c-Jun (94). Although Egr-1 is clearly nuclear in serum-stimulated or exponentially growing cells (maintained in 10% calf serum), staining of Egr-1 derivatives or fusion proteins in serum-starved cells should be examined to assess the possibility of conditional nuclear localization.

In conclusion, deletion analysis and Egr-1–β-galactosidase fusions demonstrate that nuclear localization of Egr-1 requires a bipartite signal consist-

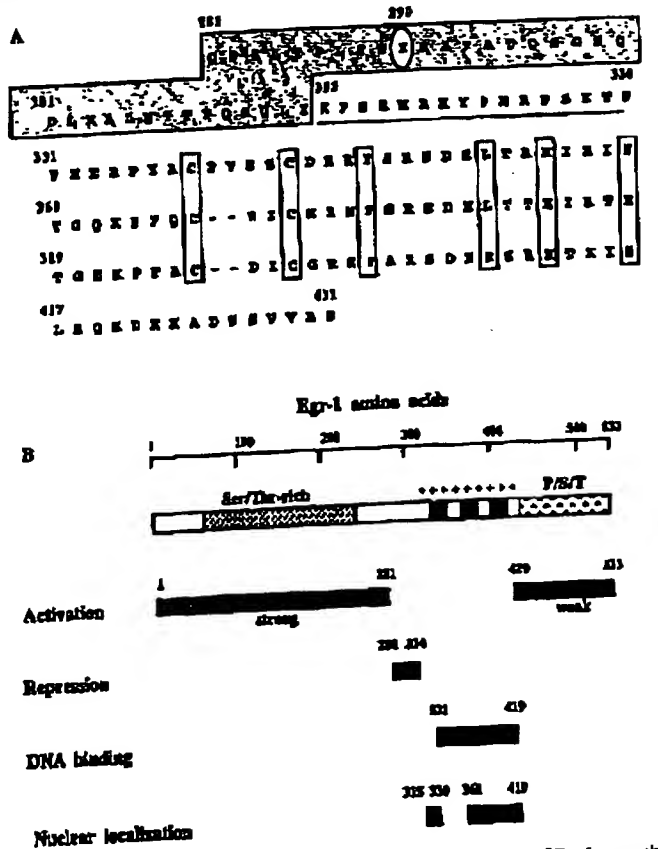


FIG. 6. Summary of Egr-1 domains (modified from 70). (A) Sequence of Egr-1 repression domain and zinc fingers. The repression domain is shaded and the 5′ basic region involved in nuclear localization is underlined. The threonine residue whose phosphorylation may prevent repression is circled. The three zinc fingers of Egr-1 are aligned for comparison, with residues conserved among Cys_2His_2 zinc fingers enclosed. (B) Functional domains of Egr-1. The serine-/threonine-rich N-terminus of Egr-1 is shown. The basic region of Egr-1 is indicated (+++). Each zinc finger is designated by a black bar and the proline/serine/threonine-rich C-terminal domain (P/S/T) is indicated. Residues 3–281 activate transcription 100-fold and the C-terminus of Egr-1 (residues 420–533) encodes a weaker *trans*-activation function. Amino acids 281–314 suffice to act as a repressor of transcription when fused to a heterologous DNA-binding domain. The DNA-binding activity of Egr-1 has been mapped to amino acids 331–419. The NLS of Egr-1 is bipartite: a basic region (amino acids 315–330) and part of the zinc-finger domain suffice to target Egr-1 to the nucleus.

ing of basic residues 315–330, which flank the zinc-finger domain, in addition to sequences within fingers 2 or 3. These results are notable in light of the fact that relatively few Cys_2His_2 zinc-finger proteins have been characterized with respect to their requirements for nuclear targeting. The incorporation of an NLS within or adjacent to the DNA-binding domain is suggestive of a conserved composite motif in Cys_2His_2 zinc-finger transcription factors (see Fig. 6). Finally, Egr-1 is a member of a small class of proteins that have bipartite nuclear localization signals in which the essential subdomains are separated by more than a few amino acids.

D. Targets of Egr-1 Regulation

1. Genes Regulated in the Context of Cellular Proliferation

Consistent with its induction by mitogenic cues and during terminal differentiation in a few cell types, Egr-1 may bind and regulate genes involved in mitosis or needed for specialized cell functions. The universality of *Egr-1* expression in response to growth signals suggests that genes downstream of *Egr-1* in the cascade governing cellular proliferation will be widely expressed. Several genes belong to this first class of Egr-1 targets, whose regulation presumably directs a cellular response to growth induction.

The expression of the thymidine kinase (*tk*) gene peaks during late G_1, after *Egr-1* induction, kinetics consistent with regulation by *Egr-1*. Enzymes such as thymidine kinase, integral to the biosynthesis of DNA, are regulated depending on the growth state of the cell, and as such thymidine kinase represents a physiologically relevant target for Egr-1. The use of specific α-Egr-1 antiserum (95) has demonstrated that Egr-1 is a component of the *tk* promoter-binding complex derived from serum-stimulated nuclear extract. Second, transient transfections in CV-1 cells show that Egr-1 activates a reporter driven by a *tk* promoter fragment from −174 to +159. Egr-1 activation appears to work through a lower affinity binding site, CCG-TCG-GTG. However, it should be noted that because *tk* is also expressed highly in actively cycling cells (in the absence of Egr-1 induction), high-level expression of *tk* apparently does not require Egr-1.

A second target for Egr-1 may be the PDGF A chain, a potent mitogen for cells of mesenchymal origin. PDGF-A is also found at high levels in a number of transformed cell lines. In normal cultured cells, levels of PDGF-A mRNA rise in response to growth factors or cytokines, but peak later than Egr-1 induction. A region of hypersensitivity to the single-strand-specific nuclease S1 in the 5′ untranslated region of PDGF-A that may be involved in regulating transcription of this growth factor has recently been defined (96).

Gel-shift competitions with purified Egr-1 showed that this homopurine/homopyrimidine site competes as well as the Egr-1 consensus. Although the S1-sensitive sequence GAG-GAG-GAC-GAGGA deviates at only the underlined position from the contacts determined by crystallography to be important for binding (64), the high affinity of the S1-sensitive site is surprising considering that previous studies have shown that GAG in the first or third subsite is not optimal for Egr-1 binding (97). Nevertheless, this homopurine/homopyrimidine sequence may be of widespread importance, because similar motifs derived from the promoters of other growth-related genes, such as the epidermal growth factor receptor, the insulin receptor, c-Ki-*ras*, c-*myc*, and TGF-β3, are also good competitors of Egr-1 binding (96). Future studies will determine if these provocative *in vitro* studies are of physiological significance by assessing whether Egr-1 can regulate transcription of the PDGF-A gene through this variant motif.

A third Egr-1 target in primary fetal astrocytes may be bFGF. An antisense oligomer to *Egr-1* blocks bFGF induction following addition of a mitogen, ET-3 (23).

2. Genes Regulated in the Context of Cell Differentiation

Expression of the myosin heavy chain α gene (*α-MHC*) and *Egr-1* are coregulated in serum-deprived primary cultures of cardiac myocytes stimulated with serum and when the embryonal carcinoma cell line P19 differentiates into cardiac cells in response to dimethyl sulfoxide, prompting investigation of α-MHC as a target of Egr-1 regulation. A CAT reporter containing 1.7 kb of the α-myosin heavy chain promoter is activated 10-fold by an *Egr-1* expression vector in transfected primary cultures of fetal rat cardiac myocytes. Northern analysis shows the endogenous *α-MHC* gene is also stimulated three- to fourfold by Egr-1 (98). Induction of *α-MHC* in response to Egr-1 was observed in the myogenic Sol8 cell line, but not in NIH3T3 fibroblasts, suggesting tissue-specific induction; *α-MHC* expression was unchanged in response to Egr-1 in another muscle cell line, L_6E_9, showing that Egr-1 is not sufficient for the *MHC* gene activation. The region of the rat *α-MHC* promoter that is Egr-1 responsive to a segment from −1698 to −1283 has been delimited (98). A potential Egr-1 binding site GTG-GGC-GTG is located within this promoter fragment, but has not yet been shown to be the functional element (98). In light of the study showing that Egr-1 translation is blocked during Sol8 differentiation in response to insulin (20), Egr-1 protein levels in cardiac myocytes remain to be analyzed.

A functional role for Egr-1 in adrenergic differentiation, suggested by high-level expression in the rat adrenal gland and in PC12 cells, may be the

regulation of phenylethanolamine *N*-methyltransferase (PNMT), the adrenal enzyme that converts norepinephrine to epinephrine. *In vivo*, neural stimulation causes an increase in Egr-1 protein in the adrenal medulla corresponding to a rise in PNMT expression in the same cell type (99). Transient transfections in the highly transfectable PC12 subline RS1 reveal that Egr-1 can modestly stimulate (fourfold) a PNMT reporter with 442 bp of 5′ sequence. This region includes two potential Egr-1 binding sites, an optimal consensus sequence at −165 and a proximal site GCG-GGC-CCG at −45. Cold competition experiments show that this 6 of 9 match to the optimal Egr-1 consensus is a weak but specific competitor (99).

It has been postulated that Egr-1 negatively regulates the widely expressed adenosine deaminase gene. ADA has a (C + C)-rich promoter, lacking TATA and CCAAT box elements, typical of classical housekeeping gene promoters. As discussed above, Egr-1 and Sp1 bind to distinct, although similarly (C + C)-rich, DNA-binding sites (97). Notably, deletion analysis of the ADA promoter reveals a *cis*-acting repressor element that maps to an Egr-1 site. Mutations that destroy Egr-1 binding but do not affect the Sp1 site in the 13 bp overlapping Egr-1/Sp1 motif GCG-TGG-GCG-GGGC result in a 1.5-fold enhancement in promoter activity. *In vitro*, Egr-1 and Sp1 protect overlapping segments of this complex 13-bp sequence. One hypothesis is that Egr-1 negatively regulates ADA transcription by competitively occupying the motif and displacing Sp1. Alternatively, Egr-1 may repress the ADA promoter by an active mechanism independent of the Sp1, also consistent with results described above. Evidence for this proposal is that even in the absence of an Sp1 site, mutation of the Egr-1 motif results in higher promoter activity. Future studies with varying ratios of Egr-1 and Sp1 expression vectors as well as experiments addressing the issue of whether the Egr-1 DNA-binding domain is sufficient for the negative regulation will be informative.

The definition of a consensus binding site for Egr-1 has propelled investigations to identify the genes that Egr-1 binds and regulates. The *tk* gene represents a physiologically relevant target for Egr-1 in the context of cell growth. The induction of the *tk* gene subsequent to the Egr-1 serum response, the ability of Egr-1 to bind to a site in the *tk* 5′ sequence, and transcriptional activation of the *tk* promoter by Egr-1 in transient transfections all support the idea that thymidine kinase is an important Egr-1 target. A second Egr-1 target may be the mitogen PDGF, because Egr-1 can bind to a site in the PDGF-A gene (*100*). Other potential Egr-1 target genes are clearly not relevant to cellular proliferation. The expression pattern of Egr-1 in the adult animal as well as its induction during terminal differentiation in some cell types suggest that Egr-1 plays a role in specialized cells that is distinct from its function during the G_0 to G_1 transition. In cardiac cells, the

endogenous α-myosin heavy chain gene or a transfected construct containing the α-*MHC* promoter is stimulated by Egr-1 (98). And in adrenal cells, Egr-1 can regulate the phenylethanolamine *N*-methyltransferase gene, supporting a role for Egr-1 in adrenergic differentiation (99). Additional targets of Egr-1 regulation in other differentiated cell types, for example, specific to osteoblasts or to macrophages, remain to be identified.

VI. *In Vivo* Role of Egr-1

The challenge remaining in current Egr-1 research is to relate correlative expression data and *in vitro* studies to a biological role for Egr-1. In a few instances, overexpression or antisense analyses have shown a phenotype for Egr-1. These studies have focused on differentiated cell types; despite the abundance of data showing Egr-1 induction by mitogenic signals, a role for Egr-1 in cell growth/division remains to be established. These phenotypic analyses are complicated by potential functional redundancy contributed by related members of the EGR family (see Section VII). With virtually identical DNA-binding domains, the expression of related family members may serve to mask a phenotype in Egr-1 loss-of-function experiments.

A clear-cut biological role for Egr-1 has been demonstrated in three systems. As discussed above, antisense oligomers preventing Egr-1 expression in myeloid cells block macrophage differentiation. Further, constitutive Egr-1 expression restricts the potential of HL-60 cells, rendering them incapable of differentiation along the granulocyte lineage (32). A second phenotype for Egr-1 involves its role as a positive regulator of astrocyte proliferation as discussed earlier (23). A third system in which Egr-1 plays a causal role involves the hypertrophic growth of cardiac myocytes in response to endothelin-1. *Egr-1*, as a gene rapidly induced by endothelin, was proposed to mediate cardiac hypertrophy. It has been definitively shown that endothelin-1-induced hypertrophic growth in adult rat cardiomyocytes, as assayed by increased protein synthesis, is blocked by oligomers complementary to the *Egr-1* message (*101*). Additional phenotypes for Egr-1 await experimentation in other systems. Perhaps the most exciting unanswered question is whether *Egr-1* functions as a cellular proto-oncogene in a manner analogous to *c-fos*.

VII. Egr-1 Is Part of a Gene Family, Including the Wilms Tumor Suppressor Gene WT1

Egr-1 shares a highly conserved domain, encoding the three zinc-finger motifs, with several other immediate-early genes as well as genes that func-

tion in unrelated contexts. Zinc-finger proteins of the type first described for TFIIIA contains invariant residues, including conserved cysteines and histidines, but also include nonconserved residues that presumably dictate the specificity of binding. *Egr-2/Krox20* (*102*, *103*), *Egr-3* (*69*), and *Egr-4/NGFI-C/pAT133* (*69*, *104*, *105*) encode proteins with zinc-finger domains virtually identical to that of Egr-1. The Egr-1 zinc-finger domain is over 95% identical to that of Egr-2 and 91% identical to that of Egr-3 at the amino-acid level. Most of the changes are conservative substitutions, and residues important in determining the sequence specificity of binding are absolutely conserved (Fig. 5A). The homology extends to adjacent basic sequences but drops abruptly outside this region. *Egr-2/Krox20* and *Egr-3* are strikingly induced by growth factors whereas *Egr-4/NGFI-C/pAT133* is more weakly inducible (*105*). The expression of *Egr-2/Krox 20*, restricted to the nervous system during mouse embryogenesis, generates a segment-specific pattern in the developing hindbrain (*69*, *103*, *106*). Importantly, disruption of *Egr-2/Krox20* by homologous recombination in the mouse results in postnatal death of the animal, with anatomical analysis showing severely reduced or absent rhombomeres 3 and 5 in the hindbrain; (*107*). Finally, *Egr-2/Krox20* brain expression is also transiently activated by electroconvulsive shock treatment, D1 dopamine receptor activation, and opiate withdrawal, in a pattern similar to that noted for *Egr-1/Zif 268* (*41*).

The Wilms tumor suppressor gene WT1, implicated in the genesis of this pediatric kidney malignancy, has four zinc fingers, three of which are highly homologous (67% identical) to the Egr-1 zinc-finger domain (*108*, *109*). The WT1 protein binds to the EGR consensus binding sequence GCG-CGG-GCG but with lower affinity. Moreover, the first finger of WT1 and the presence of KTS (an alternatively spliced variant) between fingers 3 and 4 dictate other sequence requirements for DNA binding (55). The mammalian activator Sp1 also has three related zinc fingers, with finger 2 most similar to EGR fingers 1 and 3 (*110*). The EGR family of proteins is also distantly related to MIG1, a yeast protein that responds to glucose repression (*111*). As suggested by the homology of the zinc-finger motifs, the sequences recognized by the EGR proteins, WT1, and Sp1 are related. Interestingly, flanking (A + T)-rich sequences play critical roles in target site recognition by MIG1. These flanking sequence preferences may reflect local DNA binding (*112*).

VIII. Conclusion and Future Perspectives

The genomic response of a cell to changes in its extracellular environment includes the induction of immediate-early transcription factor genes.

The most extensively characterized members of this group are the *fos*, *jun*, and *Egr* family members. Their discovery has allowed delineation of the "proximal" events from cell surface to nucleus that induce them: definition of intracellular signaling pathways and downstream promoter elements they target. More recent efforts have focused on events "distal" to transcription factor gene induction: characterization of the proteins involved, their interactions with each other, definition of the target DNA sequences to which they bind, structure–function analyses, negative regulation following induction, and other forms of cross-talk between these family members. Collectively, therefore, these investigations have enhanced our knowledge of signal transduction pathways and general mechanisms of transcriptional activation and repression, and protein-DNA interactions. The most important critical questions for future analysis involve the further identification of phenotypes, either by ectopic overexpression or by "underexpression" using dominant negative, antisense, or homologous recombination methodologies. Unfortunately, however, many phenotypes may be masked by redundant pathways. Nevertheless, a search for such systems will be critical to provide the substrate by which to characterize suitable physiological target genes for immediate-early transcription factor action.

ACKNOWLEDGMENT

This work was partially supported by NIH Grant CA49045 to V.P.S.

REFERENCES

1. D. Dernhardt, D. Edwards and C. Perlett, *BBA* 695, 83 (1986).
2. H. Herschman, *TIBS* 14, 455 (1989).
3. H. Herschman, *ARB* 60, 281 (1991).
4. L. Lau, *Curr. Opin. Cell Biol.* 3, 280 (1990).
5. V. P. Sukhatme, S. Kartha, F. Tsbeck, R. Tosh, R. Hoover and C.-H. Tsai-Morris, *Oncogene Res.* 1, 343 (1987).
6. V. Sukhatme, X. Cao, L. Chang, C.-H. Tsai-Morris, D. Stamenkovich, P. Ferreira, D. Cohen, S. Edwards, T. Shows, T. Curran, M. Le Beau and E. Adamson, *Cell* 53, 37 (1988).
7. S. Sugg, J. Kalnowitz, C. Tsai-Morris and V. Sukhatme, *NARes* 18, 4283 (1990).
8. J. Milbrandt, *Science* 238, 797 (1987).
9. L. Lau and D. Nathans, *PNAS* 84, 1182 (1987).
10. R. Liu, R. Varnum and H. Herschman, *Oncogene* 1, 263 (1987).
11. D. Simmons, D. Levy, Y. Yannoni and R. Erikson, *PNAS* 86, 1178 (1989).
12. J. Wright, K. Gunter, H. Mitsuya, S. Irving, K. Kelly and U. Siebenlist, *Science* 248, 588 (1990).
13. P. Lemaire, O. Revelant, R. Bravo and P. Charnay, *PNAS* 85, 4691 (1988).

14. B. Christy, L. Lau and D. Nathans, *PNAS* 85, 7857 (1988).
15. S. Qureshi, C. Joseph, M.-H. Rim, A. Maroney and D. Foster, *Oncogene* 6, 935 (1991).
16. K. Alexandropoulos, S. A. Qureshi, M. Rim, V. P. Sukhatme and D. A. Foster, *NARes* 20, 2353 (1992).
17. V. Seyfert, V. Sukhatme and J. Monroe. *MCBiol* 9, 2083 (1989).
18. V. Seyfert, S. McMahon and W. Glenn. *Science* 250, 797 (1990).
19. H. D. Rupprecht, P. Dunn, V. P. Sukhatme, R. B. Sterzel and D. L. Coleman, *Am. J. Physiol.* 263, F623 (1992).
20. B. Wolfrik, C. Kubisch, A. Mazza, H. Vetter and L. Neyses, *BBRC* 194, 642 (1993).
21. X. M. Cao, C. R. Gay, V. P. Sukhatme and T. H. Tan, *JBC* 267, 1345 (1992).
22. F. Rastinejad and H. Blau, *Cell* 72, 903 (1993).
23. R.-M. Hu and E. R. Levin. *J. Clin. Invest.* 93, 1826 (1994).
24. J. Milbrandt, *Neuron* 1, 183 (1988).
25. M. Watson and J. Milbrandt, *Development* 110, 173 (1990).
26. A. McMahon, J. Champion, J. McMahon and V. Sukhatme, *Development* 105, 281 (1989).
27. S. Vainio, I. Karavanova, A. Jowett and I. Thesleff, *Cell* 75, 45 (1993).
28. T. Durland, S. Samuels, V. P. Sukhatme and E. Adamson. *Oncogene* 6, 1367 (1991).
29. L. Sura, M. Ernst and G. Rodan. *MCBiol* 11, 2503 (1991).
30. S. Kharbanda, E. Rubin, R. Datta, R. Hass, V. Sukhatme and D. Kufe. *Cell Growth Differ.* 4, 17 (1993).
31. S. Kharbanda, T. Nakamura, R. Stone, R. Hass, S. Bernstein, R. Datta, V. P. Sukhatme and D. Kufe, *J. Clin. Invest.* 88, 571 (1991).
32. H. Nguyen, B. Hoffman-Liebermann and D. Liebermann, *Cell* 72, 197 (1993).
33. M. Le Beau, R. Espinosa III and W. Neuman, *PNAS* 90, 5484 (1993).
34. J. V. Bonventre, V. P. Sukhatme, M. Bamberger, A. J. Ouellette and D. Brown. *Cell Reg.* 2, 251 (1991).
35. D. Halladan, V. P. Sukhatme, M. Sherman. S. Virudachalam, D. Kufe and R. R. Weichselbaum. *PNAS* 88, 2156 (1991).
36. J. Morgan and T. Curran, *Trends Neurosci.* 12, 459 (1989).
37. D. Bartel, M. Sheng, L. Lau and M. Greenberg, *Genes Dev.* 3, 314 (1989).
38. K. Bosse, M. McCormack and L. Villa-Komaroff. *PNAS* 89, 5437 (1992).
39. P. Worley, B. Christy, Y. Nakabeppu, R. Bhat, A. Cole and J. Baraban, *PNAS* 88, 5106 (1991).
40. A. Cole, D. Saffen, J. Baraban and P. Worley, *Nature NB* 340, 474 (1989).
41. R. Bhat, P. Worley, A. Cole and J. Baraban, *Brain Res. Mol. Brain Res.* 12, 263 (1992).
42. T. Herdegen, K. Kovary, J. Leah and R. Bravo, *J. Comp. Neurol.* 313, 178 (1991).
43. C. Jamieson, R. Mayforth, M. Villereal and V. P. Sukhatme, *J. Cell Physiol.* 139, 262 (1989).
44. V. Seyfert, S. McMahon, W. Glenn, X. Cao, V. Sukhatme and J. Monroe, *J. Immunol.* 145, 3647 (1990).
45. S. A. Qureshi, M. Rim, J. Bruder, W. Kolch, U. Rapp, V. P. Sukhatme and D. A. Foster, *JBC* 266, 20594 (1991).
46. C.-H. Tsai-Morris, X. Cao and V. P. Sukhatme, *NARes* 16, 8835 (1988).
47. P. Changelian, P. Feng, T. King and J. Milbrandt, *PNAS* 88, 377 (1989).
48. B. Treisman, *Cell* 46, 567 (1986).
49. B. Christy and D. Nathans, *MCBiol* 9, 4889 (1989).
50. X. Cao, R. Koski, and A. Gashler *et al. MCBiol* 10, 1931 (1990).
51. D. Gius, X. Cao, P. Rauscher, D. Cohen, T. Curran and V. Sukhatme. *MCBiol* 10, 4243 (1990).
52. S. Qureshi, X. Cao, V. Sukhatme and D. Foster. *JBC* 266, 10802 (1991).

53. R. Datta, E. Rubin, V. Sukhatme, S. A. Qureshi, R. Weichselbaum and D. W. Kufe. *PNAS* 89, 10149 (1992).
54. J. Miller, A. McLachlan and A. Klug, *EMBO J.* 4, 1609 (1985).
55. I. Drummond, H. Rupprecht, P. Rohwer-Nutter, J. Lopez-Guisa, S. Madden, F. Rauscher III and V. Sukhatme, *MCBiol* 14, 3800 (1994).
56. J. Lanfear, T. Jowett and P. Holland, *BBRC* 178, 1229 (1991).
57. M. Russo, C. Matheny and J. Milbrandt, *MCBiol* 13, 6858 (1993).
58. R. Richards and G. Sutherland, *Cell* 70, 709 (1992).
59. J. Cordes. *TIBS* 15, 383 (1990).
60. M. Day, T. Fahrner, S. Aylsrud and J. Milbrandt, *JBC* 25, 15253 (1990).
61. C. Waters, D. Hancock and G. Evan, *Oncogene* 5, 669 (1990).
62. F. Lemaire, C. Vesque, J. Schmitt, H. Stunnenberg, R. Frank and P. Charnay, *MCBiol* 10, 3456 (1990).
63. J. Berg, *Ann. Rev Biophys. Biophys. Chem.* 19, 405 (1990).
64. N. P. Pavletich and C. O. Pabo, *Science* 252, 809 (1991).
65. E. S. Klevit, *Science* 253, 1367 (1991).
66. J. Hardelli, T. Gibson, C. Vesque and P. Charnay, *Nature NB* 343, 175 (1991).
67. T. Wilson, M. Day, T. Pentos, K. Padgett, M. Johnston and J. Milbrandt, *JBC* 267, 3715 (1992).
68. L. Fitzall, J. Schwabe, L. Chapman, J. Finch and D. Rhodes, *Nature NB* 356, 483 (1992).
69. S. Patwardhan, S. Gashler, M. Siegel, L. Chang, L. Joseph, T. Shows, M. Le Beau and V. Sukhatme, *Oncogene* 6, 917 (1991).
70. A. L. Gashler, S. Swaminathan and V. P. Sukhatme, *MCBiol* 13, 4556 (1993).
71. L. Thiell, J.-L. Castillo, D. Wu and M. Karin, *Nature NB* 342, 945 (1989).
72. N. Zelenetz-Le, A. Gashler and V. Sukhatme, unpublished (1994).
73. J. Licht, M. Crossel, J. Figge and U. Hansen, *Nature NB* 346, 76 (1990).
74. P. Zuo, D. Stanojevic, J. Colgan, K. Han, M. Levine and J. Manley, *Genes Dev.* 5, 254 (1991).
75. K. Han and J. Manley, *Genes Dev.* 7, 491 (1993).
76. S. Madden, D. Cook and F. J. Rauscher, *Oncogene* 8, 1713 (1993).
77. N. Zelenetz-Le, V. Sukhatme, R. Dropkins and D. Reinberg, unpublished (1993).
78. N. Zelenetz-Le and V. Sukhatme, unpublished (1993).
79. S.-J. Chen, E. Klann, M. Gower, C. Powell, J. Sessoms and J. Sweatt, *BJ* 28, 1032 (1990).
80. S. Hahn, *Curr. Biol.* 3, 153 (1992).
81. C. Ahnte, D. Lek and T. Curran, *MCBiol* 12, 3624 (1991).
82. J. Garcia-Bustos, J. Heitman and M. Hall, *BBA* 1071, 83 (1991).
83. P. Silver, L. Keegan and M. Ptashne, *PNAS* 81, 5951 (1984).
84. I. Treisman and I. Verma, *Oncogene* 6, 2049 (1991).
85. A. Gutchon-Maniel, P. Lescop, S. Christin-Maitre, H. Loosfelt, M. Perrot-Applanat and E. Milgrom, *EMBO J.* 10, 3551 (1991).
86. M. Hall, C. Craft and Y. Hiraoka, *PNAS* 87, 6954 (1990).
87. J. Robbins, S. Dilworth, R. Laskey and C. Dingwall, *Cell* 64, 615 (1991).
88. T. Moll, G. Tell, V. Surana, H. Robitsch and K. Nasmyth, *Cell* 66, 743 (1991).
89. J. Kleinschmidt and A. Seiler, *EMBO J.* 7, 1505 (1988).
90. S. Nath and D. Nayak, *MCBiol* 10, 4139 (1990).
91. N. Morin, C. Delsert and D. Hentz, *MCBiol* 9, 4372 (1989).
92. M. Hall, L. Hereford and I. Herskowitz, *Cell* 36, 1057 (1984).
93. H.-P. Bihs and R. Peters, *EMBO J.* 8, 1479 (1989).
94. P. Baur, J.-M. Blanchard, A. Fernandez, N. Lamb, P. Jeanteur and M. Piechaczyk, *Cell* 63, 341 (1990).

95. A. Crumt, C. Mohar and A. B. Pardee, *MCBiol* (1995). Submitted.
96. Z. Wang and T. Deuel, *BBRC* 158, 433 (1992).
97. B. Chabty and D. Nathans, *PNAS* 85, 8737 (1988).
98. M. P. Gupta, C. Gupta, R. Zak and V. P. Sukhatme, *JBC* 265, 12813 (1991).
99. S. Ebert, S. Ball, J. Hunter, A. Gubler, V. Sukhatme and D. Wong, *JBC*, 268, 20585 (1994).
100. Z. Wang, S. Madden, T. Deuel and F. Rauscher, *JBC* 267, 21999 (1992).
101. L. Neynes, J. Nouskas and H. Vetter, *BBRC* 181, 22 (1991).
102. P. Chavrier, M. Zerial, P. Lemaire, J. Almendral, B. Bravo and P. Charnay, *EMBO J.* 7, 29 (1988).
103. L. Joseph, M. Le Beau and G. Jamieson, *et al.*, *PNAS* 85, 7164 (1988).
104. S. Crosby, J. Poetz, K. Stenburger, T. Fahrner and J. Milbrandt, *MCBiol* 11, 3835 (1991).
105. H.-J. Müller, C. Skerka, A. Bialonski and P. Zipfel, *PNAS* 88, 10079 (1991).
106. D. Wilkinson, S. Bhatt, P. Chevrier, R. Bravo and P. Charnay, *Nature NB* 337, 461 (1989).
107. S. Schneider-Maunoury, P. Topilko and T. Seitandou *et al.*, *Cell*, 75, 1199 (1993).
108. M. Gessler, A. Pumtka, W. Cavenee, R. Neve, S. Orkin and C. Bruns, *Nature NB* 343, 774 (1990).
109. K. Call, T. Glaser, C. Ito, A. Buckler, J. Pelletier, D. A. Haber, E. A. Rose, A. Kral, H. Yeger, W. H. Lewis, C. Jones and D. E. Housman, *Cell* 60, 509 (1990).
110. J. Kadonaga, K. Carner, F. Masiarz and R. Tjian, *Cell* 51, 1079 (1987).
111. J. Nehlin and H. Ronne, *EMBO J.* 9, 2891 (1990).
112. M. Landis, J. Nehlin and H. Ronne, *MCBiol* 14, 1979 (1994).
113. M. H. Little, J. Prosser, A. Condie, P. J. Smith, V. Van Heyningen and N. D. Hastie, *PNAS* 89, 4791 (1992).
114. H. Rupprecht, V. P. Sukhatme, J. Lacy, H. S. Stenzel and D. L. Coleman, *Am. J. Physiol.* 253, F351 (1987).
115. D. L. Coleman, A. H. Barths, V. P. Sukhatme, J. Lin and H. D. Rupprecht, *J. Immunol.* 148, 3045 (1992).
116. R. Datta, N. Taneja, V. P. Sukhatme, S. A. Qureshi, R. Weichselbaum and D. W. Kufe, *PNAS* 89, 2419 (1992).
117. L. Neynes, J. Nouskas, J. Luyken, S. Frenhoffs, S. Oberdorf, U. Pfeifer, R. S. Williams, V. P. Sukhatme and H. Vetter. *J. Hypertens.* 11, 927 (1993).
118. D. M. Cohen and S. R. Gullans, *Am. J. Physiol.* 264, F893 (1993).
119. B. A. Haber, K. L. Mohn, R. H. Diamond and R. Taub, *J. Clin. Invest.* 91, 1319 (1993).
120. A. Sachinidis, P. Weisser, Y. Xu, K. Schulte, K. Meyer zu Brickwedde, L. Neynes and H. Vetter, *FEBS Lett.* 313, 109 (1992).
121. J. W. Liu, J. Lacy, V. P. Sukhatme and E. L. Coleman, *JBC* 268, 5629 (1991).

Two New Collagen Subgroups: Membrane-associated Collagens and Types XV and XVIII

TAINA PIHLAJANIEMI AND MARKO RIEHN

Collagen Research Unit
Biocenter and Department of Medical Biochemistry
University of Oulu
FIN-90220 Oulu, Finland

The collagens comprise a large family of genetically distinct but structurally related proteins that are found in essentially all connective tissues of most multicellular organisms, being particularly abundant in cartilage, demineralized bone, ligaments, placenta, tendon, skin, and most blood vessels. A prominent function of collagens is to maintain the architecture of tissues and organs and to confer strength on them, but they are also involved in early development and organogenesis, cell attachment, chemotaxis, and filtration through basement membranes.

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7164–7168, October 1988
Biochemistry

Molecular cloning, sequencing, and mapping of *EGR2*, a human early growth response gene encoding a protein with "zinc-binding finger" structure

(cell growth/transcriptional regulator/multigene family/DNA-binding domain)

LOREN J. JOSEPH*, MICHELLE M. LE BEAU†, GORDON A. JAMIESON, JR.‡, SONIA ACHARYA*, THOMAS B. SHOWS§, JANET D. ROWLEY†, AND VIKAS P. SUKHATME*¶

*Department of Medicine, Howard Hughes Medical Institute, †Joint Section of Hematology/Oncology, Department of Medicine, ‡Department of Pharmacological and Physiological Sciences, University of Chicago, Chicago, IL 60637; and §Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, NY 14263

Contributed by Janet D. Rowley, June 21, 1988

ABSTRACT **Early growth response gene-1 (*Egr-1*) is a mouse gene displaying *fos*-like induction kinetics in diverse cell types following mitogenic stimulation. *Egr-1* encodes a protein with "zinc-binding finger" structure. Zinc fingers are a protein structural motif that serve as DNA-binding domains in several transcriptional regulatory proteins. Using low-stringency hybridization with an Egr-1 cDNA probe, we identified a distinct human cDNA (designated EGR2 for early growth response gene-2), which is coregulated with *EGR1* by fibroblast and lymphocyte mitogens; however, several stimuli that induce Egr-1 mRNA in PC12 (rat pheochromocytoma) cells do not induce Egr-2 mRNA. The cDNA sequence predicts a protein of 406 amino acids, including three tandem zinc fingers of the Cys_2-His_2 class. Strikingly, the deduced amino acid sequences of human EGR2 and mouse Egr-1 are 92% identical in the zinc finger region but show no similarity elsewhere. *EGR2* maps to human chromosome 10 at bands q21–22. Structure–function analysis of EGR2 and EGR1 proteins should provide insight into the mechanisms linking signal transduction and transcriptional regulation of gene expression.**

Genes controlling proliferation or differentiation of eukaryotic cells have been identified by differential screening (1–5) of cDNA libraries. We (6, 7) and others (2, 3) have identified cDNAs the expression of which is upregulated by serum stimulation of quiescent mouse fibroblasts. One cDNA from our initial screening was also induced by epithelial cell and lymphocyte mitogens (7). The cDNA for this mouse early growth response gene (*Egr-1*) encodes a protein that contains three "zinc-binding fingers" of the Cys_2-His_2 subclass (6, 8, 9). *Egr-1* expression is also modulated during neuronal (4, 6) and cardiac differentiation and after cellular depolarization (6). These data suggest a role for *Egr-1* as a nuclear intermediary in signal transduction. We used low-stringency hybridization with an *Egr-1* finger-region probe to isolate several distinct human cDNAs. We report the cDNA sequence, functional characterization, and mapping of one of these clones designated EGR2.‖ Recently, others have used a similar procedure to identify murine clones that cross-hybridize with the Krüppel finger region (10–12). We show that one of these clones, Krox-20, is the murine homologue of *EGR2*. Egr-1 [NGF1-A of Milbrandt (4)], EGR2/Krox-20, and additional EGR cDNAs (L.J.J., V.P.S., unpublished data) encode zinc fingers with remarkable amino acid sequence conservation *throughout* the putative DNA-binding domains, suggesting that they recognize a similar set of target DNA sequences. The differences outside of the finger domains might be important in understanding their other regulatory interactions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Cell Culture. Cell lines 303 and HSWP (human foreskin fibroblasts) are from J. R. Smith (Baylor College of Medicine) and M. Regan (Oak Ridge), respectively. PC12 cells were provided by C. Palfrey (University of Chicago). Cell culture methods were as described (6).

RNA and Southern Hybridizations (13). All blots were done with GeneScreen*Plus* (New England Nuclear–DuPont), except RNA dot blots, for which GeneScreen was used. Hybridizations were at 65°C in 1% $NaDodSO_4$/10% dextran sulfate/1 M NaCl for 16 hr. Filters were washed at room temperature in 2× SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7), next at 65°C in 2× SSC/1% $NaDodSO_4$ (low stringency), then at 65°C in 2× SSC/1% $NaDodSO_4$ (moderate stringency), and finally at 65°C in 0.1× SSC (high stringency). Probes were made by random hexamer priming (14). RNA for hybridization analysis was isolated by the method of Chirgwin *et al.* (15), whereas for dot blots the method of Cheley and Anderson (16) was used.

DNA Sequencing. Sequencing was done by the dideoxynucleotide chain-termination method of Sanger *et al.* (17) and the double-stranded method of Zagursky *et al.* (18).

Chromosomal Localization. The methods used have been described (19–23).

RESULTS

Isolation of EGR2 cDNA Clones. A 2.1-kb *Apa* I–*Apa* I mouse *Egr-1* fragment that includes the finger region (6) was used at low stringency to screen a lambda ZAP (Stratagene, San Diego, CA) cDNA library constructed from RNA extracted from cells (303 cell line) 3 hr after serum (20%) stimulation and cycloheximide (10 μg/ml) treatment. Of several positive plaques obtained, clones Zap 2, Zap 8, and Zap 32 (Fig. 1) hybridized to a finger-region probe from *Egr-1* but not to probes flanking the finger region and contained common restriction fragments when cut with 4-base cutters. Fig. 2*A* shows an RNA blot of cell line 303 3 hr after stimulation with serum and cycloheximide probed with the

Abbreviations: *EGR1* and *EGR2*, human early growth response genes; *Egr-1* and *Egr-2*, rodent early growth response genes; nt, nucleotide(s); PMA, phorbol 12-myristate 13-acetate.

¶To whom reprint requests should be addressed.

‖The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04076).

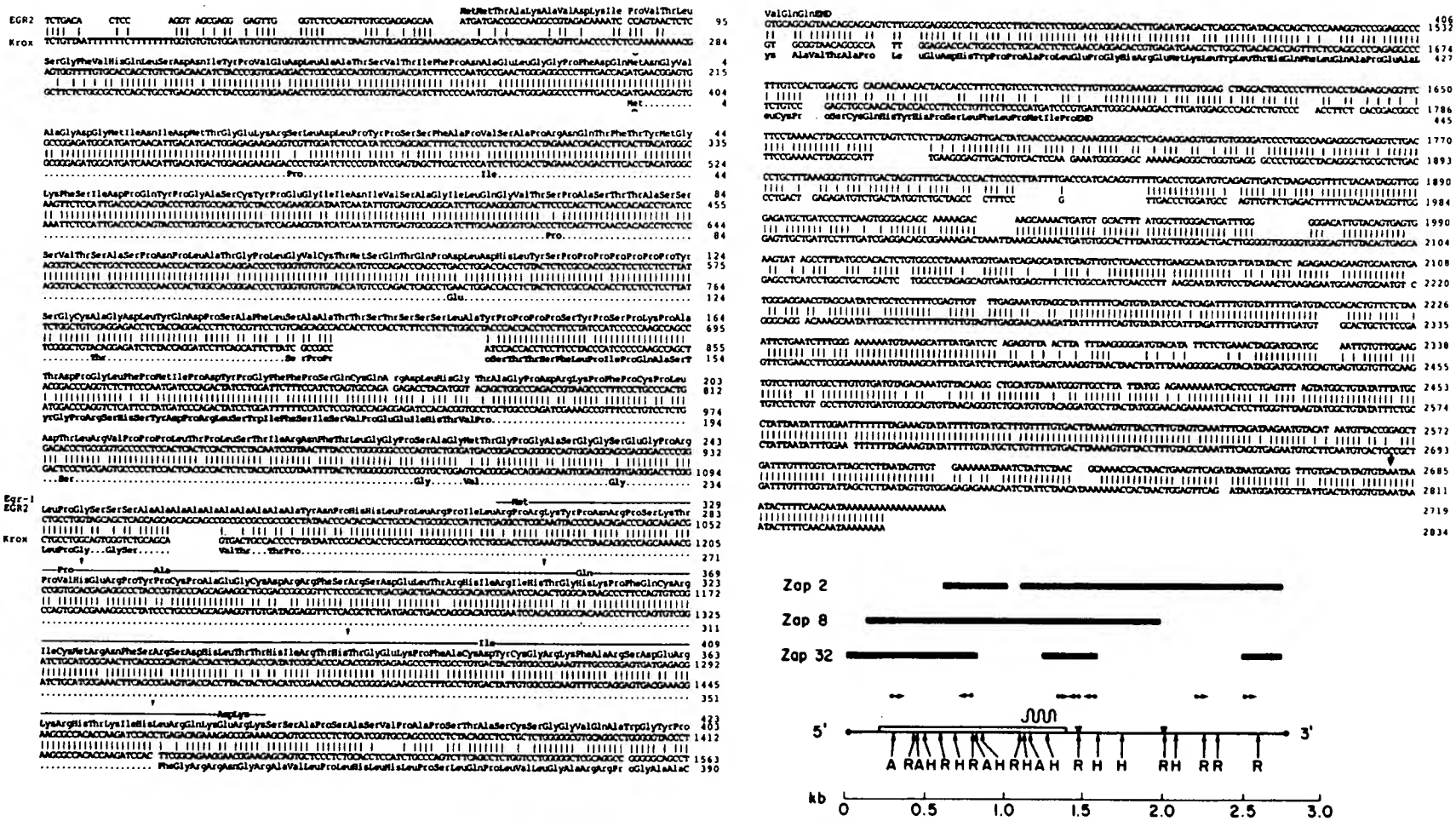


FIG. 1. Restriction map of *EGR2* and sequence comparison with Krox-20 and Egr-1 clones. The open box indicates the coding sequence of EGR2. The regions of each cDNA clone that has been sequenced are darkened. *All Rsa* I (R) restriction endonuclease sites are indicated, *only* those *Hae* III (H) and *Alu* I (A) endonuclease sites used to subclone fragments for sequencing are shown. The location of synthetic oligonucleotide primers and their orientations are shown (●→). Arrowheads (▼) indicate the boundaries of the *Rsa* I–*Rsa* I fragment subcloned into pUC19, designated P80. The zinc fingers are shown schematically. The *EGR2* nucleotide sequence is shown with the deduced amino acid sequence, including a 5′ region extending beyond the putative initiator methionine, which is marked by a carat. The Krox-20 nucleotide and amino acid sequences are shown below EGR2. The finger domain of Egr-1 is shown above the corresponding amino acid sequence of EGR2. Amino acids of Krox-20 sequence and of Egr-1 protein that are identical to the corresponding amino acid of EGR2 protein are shown as (· · ·) and (- - -), respectively. Each finger domain is marked by arrowheads. The polyadenylylation consensus signal is denoted by an arrow.

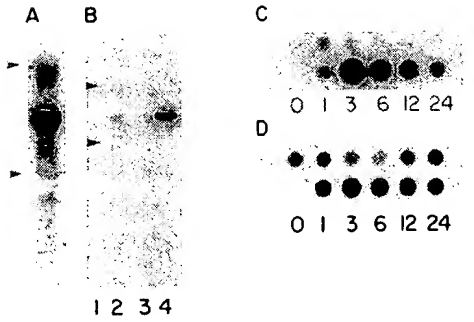


FIG. 2. EGR2 mRNA expression in human fibroblasts. (*A*) RNA blot analysis of EGR2 mRNA expression in confluent serum-deprived cell-line 303 human fibroblasts 3 hr after stimulation with 20% fetal calf serum and cycloheximide. Ten micrograms of total RNA was electrophoresed through a 1% formaldehyde gel, blotted, and probed with the ^{32}P-labeled insert of clone Zap 2. The filter was washed at moderate stringency and exposed for 18 hr without an intensifying screen. Arrowheads indicate the location of the 28S and 18S rRNAs. (*B*) RNA analysis of *EGR2* expression in confluent serum-deprived HSWP cells after stimulation with 20% fetal calf serum. Five micrograms of total RNA were loaded in each lane. Cells were harvested as follows: no treatment (lane 1); 1 hr (lane 2); 3 hr (lane 3); 3 hr after stimulation with serum and cycloheximide (lane 4). The filter was probed with the P80 insert, washed at high stringency, and exposed for 8 days with a single intensifying screen. (*C* and *D*) Extended time course of *EGR2* and *EGR1* induction. HSWP cells treated as in *B*. Replicate dot blots were prepared. Both the top row (serum only) and bottom row (serum and cycloheximide) are from cells harvested at 0, 1, 3, 6, 12, and 24 hr after stimulation. The filter in Fig. 2*C* was probed with the P80 insert. The filter in *D* was probed with an *Msp* I–*Rsa* I 700-base pairs (bp) probe from the region 3′ to the finger domains of *Egr-1* (6). The filters in *C* and *D* were washed to high stringency and exposed overnight with a single intensifying screen.

Zap 2 insert. Multiple bands were detected after moderately stringent washing, but the band at 3.2 kilobases (kb) was dominant and designated as the *EGR2*-encoded transcript.

EGR2 mRNA Is Induced in Serum-Stimulated G_0 Human Fibroblasts. We asked whether *EGR2* is induced in human fibroblasts by serum in the absence of cycloheximide. HSWP cells were used to exploit the extensive characterization of mitogen-stimulated events in that line (24, 25). Because EGR1 mRNA might cross-hybridize to *EGR2*, a *nonfinger*-encoding *Rsa* I fragment from Zap 8 (Fig. 1) was subcloned into pUC19 and designated P80. Fig. 2*B* shows that the EGR2 mRNA level is elevated in HSWP cells at 1 hr and barely detectable at 3 hr after serum stimulation. Cycloheximide addition results in superinduction. Fig. 2 *C* and *D* show a replicate dot blot comparison of *EGR2* and *EGR1* induction in these cells over 24 hr. The signal intensities suggest that the level of EGR2 mRNA induced was several-fold lower than the level of EGR1 mRNA.

EGR2 Is Induced in Phorbol 12-Myristate 13-Acetate (PMA)-Stimulated Human Mononuclear Cells. To see whether induction of EGR2 mRNA was specific for fibroblasts or a more general phenomenon, we examined human lymphocytes. Fig. 3*A* shows an RNA blot analysis of human peripheral blood mononuclear cells subsequent to PMA stimulation. The P80 probe detects two transcripts at 2.5 and 3.2 kb after high-stringency washing.

Egr-1 But Not Egr-2 mRNA Is Inducible in PC12 Cells. Several stimuli induce Egr-1 mRNA in PC12 cells (4, 6). RNA was prepared from PC12 cells 1 hr after stimulation with the agents indicated in Fig. 3*B–D*. Fig. 3*B* shows the results of hybridization with an *Egr-1* probe that includes the finger-encoding region. Fig. 3*C* shows a replicate filter probed with the Zap 2 insert that includes the finger region of *EGR2*. The filter shown in Fig. 3*D* was probed with P80, a nonfinger

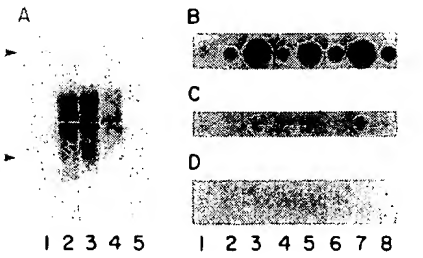


FIG. 3. EGR2 mRNA expression in human lymphocytes and in PC12 cells. (*A*) Peripheral blood was separated on Ficoll/Hypaque. Mononuclear cells (5×10^7) were used for each point. Cells were resuspended in medium containing 10% fetal calf serum and PMA (25 ng/ml). Five micrograms of total RNA was loaded in each lane. Cells were harvested at intervals after stimulation: 0 hr, no treatment (lane 1); 45 min (lane 2); 2.5 hr (lane 3); 5 hr (lane 4); 1 hr after maintenance in autologous serum (L.J.J.) without PMA (lane 5). The filter was probed with the P80 plasmid, washed to high stringency, and exposed overnight with a single intensifying screen. (*B–D*) RNA dot blot comparison of *Egr-1* and *Egr-2* expression in PC12 cells. Cells were harvested at 1 hr after stimulation with the agents indicated. Replicate RNA dot blot filters were made. (*A–D*) 1, no treatment; 2, sham treatment with serum-free medium; 3, PMA (100 nM); 4, nerve growth factor (100 ng/ml); 5, epidermal growth factor (100 ng/ml); 6, calcium ionophore A23187 (5 μg/ml); 7, bradykinin (100 ng/ml); 8, A23187 (5 μg/ml) plus PMA (100 nM). (*B*) This filter was probed with a 2.1-kb *Apa* I–*Apa* I probe, including the finger region, of *Egr-1* (6), washed to high stringency, and exposed overnight with a single intensifying screen. (*C*) This replicate was probed with the Zap 2 insert, washed to moderate stringency, and exposed overnight with a single screen. (*D*) This filter was probed with the insert of P80, washed to moderate stringency, and exposed overnight with a single screen.

region probe from *EGR2*: no hybridization is detected after a 3-day exposure. These results indicate that the weak signal seen in Fig. 3*C* is from cross-hybridization due to the finger region and sets an upper limit on the level of induction of other zinc finger-encoding transcripts.

cDNA Sequence. Fig. 1 shows the restriction map of the EGR2 cDNA and its complete nucleotide sequence. The sequence is 2719 nucleotides (nt) long and terminates in a poly(A) tract. There are in-frame termination codons at nt 3 and 15. Following these are several methionine codons; however, none fulfill the Kozak (26) criterion for an initiator codon: RNNATGG, where R represents adenine or guanine. The most 5′ ATG is usually the functionally important initiator; a common exception is for protooncogenes (26). The methionine designated as amino acid number 1 (nt 204) for EGR2 corresponds to the most 5′ methionine reported in the Krox-20 protein. We made this choice based on the fact that the nucleotide comparison (see below) of *EGR2* and Krox-20 sequences suggests that they are homologues. The two nucleotide sequences diverge before this methionine. This ATG initiates an open reading frame of 1218 nt, terminating at the stop codon at nt 1422. A polyadenylylation signal consensus sequence, AATAAA (27), is located at nt 2681, 14 nt before the poly(A) tract.

Structural Features of the Deduced Amino Acid Sequence. The cDNA sequence predicts a protein of 406 amino acids with a M_r of 43,307. Amino acids 286–370 form three tandem zinc fingers of the form Thr-Gly-Xaa$_3$-(Tyr/Phe)-Xaa-Cys-Xaa$_{2-4}$-Cys-Xaa$_3$-Phe-Xaa$_5$-Leu-Xaa$_2$-His-Xaa$_3$-His described as a consensus sequence for members of the Cys$_2$-His$_2$ class (6, 28, 29). The fingers are connected by H–C links [Thr-Gly-Glu-(Arg/Lys)-Pro-(Phe/Tyr)-Xaa], a highly conserved motif described by Schuh *et al.* (30) and found in the Egr-1 protein and other, but not all, members of the Cys$_2$-His$_2$ family. EGR2, like Egr-1, is rich in proline (15%), serine (11%), alanine (8%), and threonine (7%) residues. There is a

Table 1. Concordancy analysis of somatic cell hybrid panel

	Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
Concordant	(+/+)	7	10	16	10	16	9	12	15	8	25	11	19	10	18	11	9	17	16	6	18	18	6	11
hybrids, no.	(−/−)	7	6	6	6	6	7	3	4	7	7	6	5	4	3	7	7	3	6	6	5	3	5	4
Discordant	(+/−)	16	16	7	16	10	17	13	11	16	0	13	7	16	8	15	17	7	10	20	8	8	19	8
hybrids, no.	(−/+)	0	1	1	1	1	0	3	3	0	0	1	2	3	4	0	0	4	1	1	2	4	2	3
	% Discordancy	53	52	27	52	33	52	52	42	52	0	45	27	58	36	45	52	35	33	64	30	36	66	42

The table is compiled from 33 cell hybrids involving 15 unrelated human cell lines and 4 mouse cell lines (21–23). The P80 *EGR2* probe was hybridized to Southern blots of *Eco*RI-digested DNA from human–mouse hybrids. The *EGR2* gene localization was determined by scoring the presence or absence of a human band in the hybrids. Concordant hybrids have either retained or lost the human *EGR2* band together with a specific chromosome or the reverse. These concordances are designated (+/+) and (−/−), respectively, where the first symbol denotes the presence or absence of the human *EGR2* band and the second symbol denotes the presence or absence of the specific human chromosome. Percent discordancy indicates the degree of discordant segregation for a marker and a chromosome ([−/+] or [+/−]). A 0% discordancy is the basis for chromosome assignment. The *EGR2* gene mapped to human chromosome 10. DNA from hybrid XTR-3BSAgB with no intact chromosome 10 but retaining 10pter → 10q23 showed hybridization with the P80 probe indicating that *EGR2* is located 10pter → 10q23.

run of 7 consecutive prolines (amino acids 117–123) and of 13 serines and alanines (amino acids 247–259). The high content of threonines and serines suggests that EGR2 could be phosphorylated, a potentially important means of regulation.

Comparison of the Amino Acid Sequence of the Zinc Finger Regions. The amino acid sequence of the finger domains of EGR2 is 100% identical with that of Krox-20 sequence (Fig. 1). The finger region sequence of EGR2/Krox-20 shows an average identity of 37% with the Krüppel fingers (11, 12, 30), due primarily to the conserved H–C link sequence. The loop of the first finger of EGR2/Krox-20 sequence matches the loop of the second finger of the transcription factor Sp1 at 8 of 12 amino acids and at 10 of 12 amino acids when one includes conservative changes (12, 29). More striking is the 92% identity between the amino acid sequence of the EGR2/Krox-20 zinc finger region and the corresponding region in Egr-1 (nucleotide identity in region is 78%). Notably, none of the four amino acid differences between Egr-1 and EGR2 are located at the "finger tips," which are thought to contact the target DNA sequence (31). There is marked sequence similarity among Egr-1, EGR2, and Krox-20 sequence immediately 5′ of the finger region and for Egr-1 and EGR2 immediately 3′ of the finger region (Fig. 1). There is no significant similarity elsewhere between Egr-1 and EGR2.

***EGR2* (Human) and Krox-20 (Murine) Sequences Are Homologues.** There is extensive nucleotide similarity between EGR2 and Krox-20 cDNAs (Fig. 1), suggesting that these two cDNAs are homologues. The overall nucleotide identity is 75% (87% in the coding region and 89% in the finger region). The amino acid identity is 84% from the initiator methionine to the last amino acid of the finger domains (after which the two deduced sequences diverge markedly because of a single nucleotide difference at position 1314 in *EGR2*—see *Discussion*).

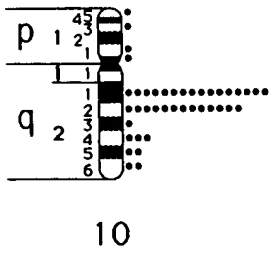


FIG. 4. Distribution of labeled sites on chromosome 10. The figure summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized with the EGR2 Zap 32 cDNA probe. Each dot indicates one labeled site observed in the corresponding band. Seventy percent (29/41) of the labeled sites on chromosome 10 were located at q21–22; this cluster represented 18.1% of all labeled sites (29/160).

Chromosomal Localization. The results of Southern blot analysis of genomic DNA from mouse × human hybrids probed with the P80 (nonfinger region) plasmid are shown in Table 1. The discordancy scores localize the *EGR2* gene to chromosome 10. To determine the chromosomal sublocalization of *EGR2*, we hybridized the Zap 32 plasmid to normal human metaphase chromosomes. Of 100 metaphase cells examined from this hybridization, 31 (31%) were labeled on region q2 of one or both chromosome 10 homologues. The distribution of labeled sites on chromosome 10 is illustrated in Fig. 4; of 160 total labeled sites observed, 41 (25.6%) were located on this chromosome. These sites were clustered at bands q21–22, and this cluster represented 18.1% (29/160) of all labeled sites (cumulative probability for the Poisson distribution is $<<0.0005$). Thus, these results indicate that the *EGR2* gene is localized to chromosome 10 at bands q21–22. However, 20 grains, representing 12.5% ($P < 0.0005$) of all labeled sites were seen at 3p24–26. Similar results were obtained in three additional hybridization experiments with this probe. The observation of specific labeling on both chromosomes 3 and 10 in hybridizations using the Zap 32 cDNA probe, which contains the *EGR2*-encoded finger domain sequence, raised the possibility that this probe was hybridizing to another finger domain-containing gene located on the short arm of chromosome 3. Use of the P80 probe, which does not contain finger domain sequences, resulted in specific labeling of the proximal long arm of chromosome 10. Of 147 labeled sites seen in 100 metaphase cells, 21 (14.3%, $P << 0.0005$) were located at 10q21–22. Hybridization of this probe resulted in a substantial reduction of labeling on 3p; however, a few grains of unknown significance were noted at 3p24–26. Two additional experiments resulted in specific labeling only of the long arm of chromosome 10.

DISCUSSION

A major goal of cell biology is to analyze the molecular mechanisms controlling gene expression. An important component of this network are DNA-binding proteins with transcriptional activity. The importance of such proteins in cell growth is suggested by the finding that *c-jun* (the cellular homologue of the oncogene *v-jun*) encodes the transcription factor AP1 (32, 33), the discovery of *c-jun*-related genes (34) that are growth-factor regulated, and the identification of a zinc-finger-encoding gene (*Egr-1*) coregulated with *c-fos* (6). This paper reports the isolation and characterization of a second mitogen-inducible cDNA (designated EGR2), which also encodes a protein with zinc fingers. Furthermore, its levels are elevated after growth stimulation in fibroblasts as well as in lymphocytes; thus, like *Egr-1*, the expression of *EGR2* is not restricted to one cell type. *EGR2* is therefore probably involved in the network of genes controlling the proliferative response. Whether *EGR2* acts to transmit, am-

plify, or limit responses to such stimuli is unknown. Whether *EGR2* expression is specific to the G_o–G_1 transition remains to be determined. However, unlike *Egr-1*, *Egr-2* induction is not seen in PC12 cells after stimulation by various agents, suggesting that differences exist in the 5′ regulatory regions of these genes.

Several reports show that the zinc-finger region alone confers sequence specificity of binding (35–37). The amino acids at the tips of the Cys_2-His_2 loops are thought to be responsible for DNA contact (31). The surprisingly high degree of amino acid similarity of Egr-1 and EGR2 throughout the finger region and dissimilarity elsewhere offers a rare example among the Cys_2-His_2 zinc-finger proteins for comparing structure and function. As a working hypothesis it seems reasonable that Egr-1 and EGR2 might recognize the same DNA target sequences through their zinc fingers but that interactions with other transcriptional regulatory elements could differ greatly.

The high level of nucleotide similarity throughout EGR2 and Krox-20 sequence suggests they are homologues. Although some differences in amino acid sequence could represent alternative splicing of small exons or evolutionary divergence, this possibility is unlikely to explain the extensive amino acid dissimilarity 3′ to the finger domains resulting from the single nucleotide frameshift at position 1314 in the EGR2 sequence. We sequenced this area on three independently selected clones (Fig. 1). In addition, in our predicted sequence the four amino acids immediately after the last histidine of the third zinc finger match perfectly the corresponding four amino acids in Egr-1 thereby extending the region of their amino acid identity.

The results of *in situ* chromosomal hybridization and the Southern blot analysis of somatic cell hybrids demonstrate that *EGR2* maps to 10q21–22. Although few structural rearrangements involving the long arm of chromosome 10 have been seen in human tumors, a loss of an entire chromosome 10 has been reported as a recurring abnormality in gliomas in adults (38). Relatively few genes have been mapped to 10q21–22; of these, only the gene(s) implicated in multiple endocrine neoplasia type 2A and the gene for lipocortin IIc are potentially involved in cell activation or growth (39, 40). As a result of the putative regulatory activity of *EGR2*, loss of this gene could lead to deregulated cell growth.

Other EGR cDNAs exist that encode proteins with highly related zinc fingers to those in Egr-1 and EGR2 (L.J.J., V.P.S., unpublished data). This multigene family offers a rich opportunity to investigate the relationship of signal transduction to gene expression in normal and transformed cells.

We thank L. Laimins and F. G. Toback for their critical reading of the manuscript, D. Stamenkovich, R. Espinosa, III, and R. Eddy for their expert technical assistance. L.J.J. and V.P.S. are supported by the Howard Hughes Medical Institute; M.M.L. is a Scholar of the Leukemia Society of America. G.A.J. is supported by Public Health Service Grant GM28359 (M. L. Villereal). This work was also supported by Public Health Service Grant CA42557 and U.S. Department of Energy Contract DE-FG02-86ER60408 (J.D.R.); Public Health Service Grants GM20454 and HD05196; and American Cancer Society Grant ACS-CD62 (T.B.S.).

1. Cochran, B. H., Reffel, A. C. & Stiles, C. D. (1983) *Cell* **33**, 939–947.
2. Lau, L. F. & Nathans, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1182–1186.
3. Almendral, J. M., Sommer, D., MacDonald-Bravo, H., Burckhardt, J., Perera, J. & Bravo, R. (1988) *Mol. Cell. Biol.* **8**, 2140–2148.
4. Milbrandt, J. (1987) *Science* **238**, 797–799.
5. Lim, R., Varnum, B. C. & Herschman, H. (1987) *Oncogene* **1**, 263–270.
6. Sukhatme, V. P., Cao, X., Chang, L. C., Tsai-Morris, C., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., Le Beau, M. M. & Adamson, E. D. (1988) *Cell* **56**, 337–343.
7. Sukhatme, V. P., Kartha, S., Toback, F. G., Taub, R., Hoover, R. G. & Tsai-Morris, C. (1987) *Oncogene Res.* **1**, 343–355.
8. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469.
9. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3.
10. Chowdhury, K., Deutsch, U. & Gruss, P. (1987) *Cell* **48**, 771–778.
11. Chavrier, P., Lemaire, P., Revelant, O., Bravo, R. & Charnay, P. (1988) *Mol. Cell. Biol.* **8**, 1319–1326.
12. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. & Charnay, P. (1988) *EMBO J.* **7**, 29–35.
13. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
14. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
15. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
16. Cheley, S. & Anderson, R. (1984) *Anal. Biochem.* **137**, 9–15.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Zagursky, R., Baumeister, N., Lomax, N. & Berman, M. (1985) *Gene Anal. Tech.* **2**, 89–94.
19. Shows, T. B., Sakaguchi, A. Y. & Naylor, S. L. (1982) *Advances in Human Genetics*, eds. Haris, H. & Hirschhorn, K. (Plenum, New York and London), Vol. 12, pp. 341–452.
20. Shows, T., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H. & Taniguchi, T. (1984) *Somatic Cell Mol. Genet.* **10**, 315–318.
21. Shows, T. B., Brown, J. A., Haley, L. L., Byers, M. G., Eddy, R. L., Cooper, E. S. & Goggin, A. P. (1978) *Cytogenet. Cell Genet.* **21**, 99–104.
22. Shows, T. B. (1983) *Isozymes: Current Topics in Biological and Medical Research*, eds. Rattazzi, M. C., Sandalios, J. G. & Whitt, G. S. (Liss, New York), Vol. 10, pp. 323–339.
23. Le Beau, M. M., Westbrook, C. A., Diaz, M. O. & Rowley, J. D. (1984) *Nature (London)* **312**, 70–71.
24. Muldoon, L. L., Jamieson, G. A., Jr., Kao, A. C., Palfrey, H. C. & Villereal, M. L. (1987) *Am. J. Physiol.* **253**, C219–C229.
25. Jamieson, G. A., Jr., Etschied, B. G., Muldoon, L. L. & Villereal, M. L. (1988) *J. Cell Physiol.* **134**, 220–228.
26. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8131.
27. Wickens, M. & Stefenson, P. (1984) *Science* **226**, 1045–1051.
28. Brown, R. S. & Argos, P. (1985) *Nature (London)* **324**, 215.
29. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51**, 1079–1090.
30. Schuh, R., Aicher, W., Gaul, U., Côté, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schröder, C., Kemler, R. & Jäckle, H. (1986) *Cell* **47**, 1025–1032.
31. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 99–102.
32. Angel, P., Allegretto, E., Okino, S., Hattori, K., Boyle, W. J., Hunter, T. & Karin, M. (1988) *Nature (London)* **332**, 166–171.
33. Bohmann, D., Bos, T. J., Admon, A., Tetsuji, N., Vogt, P. K. & Tjian, R. (1987) *Science* **238**, 1386–1392.
34. Ryder, K., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1487–1491.
35. Green, S. & Chambon, P. (1987) *Nature (London)* **325**, 75–78.
36. Vrana, K. E., Churchill, M. E. A., Tullius, T. D. & Brown, D. D. (1988) *Mol. Cell. Biol.* **8**, 1684–1695.
37. Nagai, K., Nakaseko, Y., Nasmyth, K. & Rhodes, D. (1988) *Nature (London)* **332**, 284–286.
38. Bigner, S. H., Mark, J., Bullard, D. E., Mahaley, M. S. & Bigner, D. D. (1986) *Cancer Genet. Cytogenet.* **22**, 121–135.
39. Simpson, N. E., Kidd, K. K., Goodfellow, P. J., McDermid, H., Myers, S., Kidd, J. R., Jackson, C. E., Duncan, A. M. V., Farrer, L. A., Brasch, K., Castiglione, C., Genel, M., Gertner, J., Greenberg, C. R., Gusella, J. F., Holden, J. J. A. & White, B. N. (1988) *Cytogenet. Cell Genet.* **46** (1–4), 693 (abstr.).
40. Huebner, K., Cannizzaro, L. A., Croce, C. M., Frey, A. Z., Wallner, B. P., Hecht, B. K. & Hecht, F. (1988) *Cytogenet. Cell Genet.* **46** (1–4), 631 (abstr.).

Correction. In the article "Molecular cloning, sequencing, and mapping of *EGR2*, a human early growth response gene encoding a protein with 'zinc-binding finger' structure" by Loren J. Joseph, Michelle M. LeBeau, Gordon A. Jamieson, Jr., Sonia Acharya, Thomas B. Shows, Janet D. Rowley, and Vikas P. Sukhatme, which appeared in number 19, October 1988, of *Proc. Natl. Acad. Sci. USA* (**85**, 7164–7168), the authors request that the following correction be noted. Sequencing of a human genomic EGR2 clone in their laboratory has identified an error in the 3'-coding region of the EGR2 cDNA shown in Fig. 1 of that article. The GG at positions 1397–1398 should be replaced by a single guanine. The resultant corrected amino acid sequence after the Gln (Q) at position 398 should read (in one-letter code) P-G-G-T-L-C-S-S-N-S-S-S-L-G-G-G-P-L-A-P-C-S-S-R-T-R-T-P. The corrected M_r of the deduced 426 amino acid protein is 45,007.

cDNA sequence of the human cellular early growth response gene Egr-1

S.V.Suggs, J.L.Katzowitz, C.Tsai-Morris[1] and V.P.Sukhatme*[1]
Amgen, Thousand Oaks, CA 91320 and [1]The University of Chicago, Chicago, IL 60637, USA

Submitted April 13, 1990 EMBL accession no. X52541

The Egr family of cellular proteins is composed of four distinct members which contain closely related Cys_2-His_2 zinc finger motifs. Egr-1 (1, 2) [zif 268 (3), NGFIA (4), Krox 24 (5) and TIS 8(6)] is induced by diverse signals that initiate mitogenesis, differentiation and by neuronal excitation (2, 3, 4, 6). Egr-1 activates transcription via the sequence CGCCCCCGC (7, 8, unpublished observations), thus suggesting that it may play a broad role as a nuclear signal transducer. The human Egr-1 gene is located on chromosome 5q23−31 (2). Since interstitial deletions in this area occur frequently in patients with therapy related acute myelocytic leukemia (t-AML) (2) and since Egr-1 levels are high in differentiated myeloid cells (unpublished observations), it is possible that Egr-1 has a role in myeloid differentiation. We have therefore isolated and fully sequenced the human Egr-1 cDNA from a human 303 fibroblast library obtained by cross-hybridization to the murine Egr-1 clone OC3.1 (2). The overall nucleotide similarity between mouse and human sequences is 87% and 94% at the nucleotide and protein levels, respectively. The protein is particularly rich in proline, serine, and threonine residues that have been found in abundance in other transcription factors. Interestingly, the zinc finger domain (a.a. 340−419) critical in DNA binding is identical in mouse (2, 3, 5), rat (4), chicken (9) and man.

REFERENCES

1. Sukhatme,V.P. *et al.* (1987) *Oncogene Res.* **1**, 343−355.
2. Sukhatme,V.P. *et al.* (1988) *Cell* **53**, 37−43.
3. Christy,B.A. *et al.* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7857−7861.
4. Milbrandt,J. (1987) *Science* **237**, 797−799.
5. Lemaire,P. *et al.* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4691−4695.
6. Lim,R.W. *et al.* (1987) *Oncogene* **1**, 263−270.
7. Christy,B.A. and Nathans,D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8737−8741.
8. Cao,X. *et al.* (1990) *Mol. Cell. Biol.* In press.
9. Simmons,D., personal communication.

```
CCGCAGAACTTGGGGAGCCGCCGCCGCCATCCGCCGCCGCAGCCAGCTTCCGCCGCCGCAGGACCGGCCCCTGCCCCAGCCTCCGCAGCCGCGGCGCGTCCACGCCCGCCCGCGCCCAGG  120
GCGAGTCGGGGTCGCCGCCTGCACGCTTCTCAGTGTTCCCCGCGCCCCCGCATGTAACCCGGCCAGGCCCCCGCAACGGTGTCCCCTGCAGCTCCAGCCCCGGGCTGCACCCCCCCGCCCC  240
                                M  A  A  A  K  A  E  M  Q  L  M  S  P  L  Q  I  S  D  P  F  G  S  F  P  H  S  P  T  M  D    30
GACACCAGCTCTCCAGCCTGCTCGTCCAGGATGGCCGCGGCCAAGGCCGAGATGCAGCTGATGTCCCCGCTGCAGATCTCTGACCCGTTCGGATCCTTTCCTCACTCGCCCACCATGGAC  360
 N  Y  P  K  L  E  E  M  M  L  L  S  N  G  A  P  Q  F  L  G  A  A  G  A  P  E  G  S  G  S  N  S  S  S  S  S  S  G  G  G    70
AACTACCCTAAGCTGGAGGAGATGATGCTGCTGAGCAACGGGGCTCCCCAGTTCCTCGGCGCCGCCGGGGCCCCAGAGGGCAGCGGCAGCAACAGCAGCAGCAGCAGCGGGGCGGT  480
 G  G  G  G  G  G  S  N  S  S  S  S  S  S  T  F  H  P  Q  A  D  T  G  E  Q  P  Y  E  H  L  T  A  E  S  F  P  D  I  S  L  110
GGAGGCGGCGGGGGCGGCAGCAACAGCAGCAGCAGCAGCAGCACCTTCAACCCTCAGGCGGACACGGGCGAGCAGCCCTACGAGCACCTGACCGCAGAGTCTTTTCCTGACATCTCTCTG  600
 N  N  E  K  V  L  V  E  T  S  Y  P  S  Q  T  T  R  L  P  P  I  T  Y  T  G  R  F  S  L  E  P  A  P  N  S  G  N  T  L  W  150
AACAACGAGAAGGTGCTGGTGGAGACCAGTTACCCCAGCCAAACCACTCGACTGCCCCCCATCACCTATACTGGCCGCTTTTCCCTGGAGCCTGCACCCAACAGTGGCAACACCTTGTGG  720
 P  E  P  L  F  S  L  V  S  G  L  V  S  M  T  N  P  P  A  S  S  S  S  A  P  S  P  A  A  S  S  A  S  A  S  Q  S  P  F  L  190
CCCGAGCCCCCTCTTCAGCTTGGTCAGTGGCCTAGTGAGCATGACCAACCCACCGGCCTCCTCGTCCTCAGCACCATCTCCAGCGGCCTCCTCCGCCTCCGCCTCCCAGAGCCCACCCCTG  840
 S  C  A  V  P  S  N  D  S  S  P  I  Y  S  A  A  P  T  F  P  T  P  N  T  D  I  F  P  E  P  Q  S  Q  A  F  P  G  S  A  G  230
AGCTGCGCAGTGCCATCCAACGACAGCAGTCCCATTTACTCAGCGGCACCCACCTTCCCCACGCCGAACACTGACATTTTCCCTGAGCCACAAAGCCAGGCCTTCCCGGGGCTCGGCAGGG  960
 T  A  L  Q  Y  P  P  P  A  Y  P  A  A  K  G  G  F  Q  V  P  M  I  P  D  Y  L  F  P  Q  Q  Q  G  D  L  G  L  G  T  P  D  270
ACAGCGCTCCAGTACCCGCCTCCTGCCTACCCTGCCGCCAAGGGTGGCTTCCAGGTTCCCATGATCCCCGACTACCTGTTTCCACAGCAGCAGGGGGATCTGGGCCTGGGCACCCCAGAC  1080
 Q  K  P  F  Q  G  L  E  S  R  T  Q  Q  P  S  L  T  P  L  S  T  I  K  A  F  A  T  Q  S  G  S  Q  D  L  K  A  L  N  T  S  310
CAGAAGCCCTTCCAGGGCCTGGAGAGCCGCACCCAGCAGCCTTCGCTAACCCCTCTGTCTACTATTAAGGCCTTTGCCACTCAGTCGGGCTCCCAGGACCTGAAGGCCCTCAATACCAGC  1200
 Y  Q  S  Q  L  I  K  P  S  R  M  R  K  Y  P  N  R  P  S  K  T  P  P  E  E  R  P  Y  A  C  P  V  E  S  C  D  R  R  F  S  350
TACCAGTCCCAGCTCATCAAACCCAGCCGCATGCGCAAGTATCCCAACCGGCCCAGCAAGACGCCCCCCCACGAACGCCCTTACGCTTGCCCAGTGGAGTCCTGTGATCGCCGCTTCTCC  1320
 R  S  D  E  L  T  R  H  I  R  I  H  T  G  Q  K  P  F  Q  C  R  I  C  M  R  N  F  S  R  S  D  H  L  T  T  H  I  R  T  H  390
CGCTCCGACGAGCTCACCCGCCACATCCGCATCCACACAGGCCAGAAGCCCTTCCAGTGCCGCATCTGCATGCGCAACTTCAGCCGCAGCGACCACCTCACCACCCACATCCGCACCCAC  1440
 T  G  E  K  P  F  A  C  D  I  C  G  R  K  F  A  R  S  D  E  R  K  R  H  T  K  I  H  L  R  Q  K  D  K  K  A  D  K  S  V  430
ACAGGCGAAAAGCCCTTCGCCTGCGACATCTGTGGAAAGAAAGTTTGCCAGGAGCGATGAACGCAAGAGGCATACCAAGATCCACTTGCGGCAGAAGGACAAGAAAGCAGACAAAAGTGTT  1560
 V  A  S  S  A  T  S  S  L  S  S  Y  P  S  P  V  A  T  S  Y  P  S  P  V  T  T  S  Y  P  S  P  A  T  T  S  Y  P  S  P  V  470
GTGGCCTCTTCGGCCACCTCCTCTCTCTCTTCCTACCCGTCCCCGGTTGCTACCTCTTACCCGTCCCCGGTTACTACCTCTTATCCATCCCCGGCCACCACCTCATACCCATCCCCTGTG  1680
 P  T  S  F  S  S  P  G  S  S  T  Y  P  S  P  V  H  S  G  F  P  S  P  S  V  A  T  T  Y  S  S  V  P  P  A  F  P  A  Q  V  510
CCCACCTCCTTCTCCTCTCCCGGCTCCTCGACCTACCCATCCCCTGTGCACAGTGGCTTCCCCTCCCCGTCGGTGGCCACCACGTACTCCTCTGTTCCCCCTGCTTTCCCGGCCCAGGTC  1800
 S  S  F  P  S  S  A  V  T  N  S  F  S  A  S  T  G  L  S  D  M  T  A  T  F  S  P  R  T  I  E  I  C                    543
AGCAGCTTCCCTTCCTCAGCTGTCACCAACTCCTTCAGCGCCTCCACAGGGCTTTCGGACATGACAGCAACCTTTTCTCCCAGGACAATTGAAATTTGCTAAAGGGAAAGGGGAAAGAAA  1920
GGGAAAAGGGAGAAAAGAAACACAAGAGACTTAAAGGACAGGAGGAGGAGATGGCCATAGGAGAGGAGGGTTCCTCTTAGGTCAGATGGAGGTTCTCAGAGCCAAGTCCTCCCTCTCTA  2040
CTGGAGTGGAAGGTCTATTGGCCAACAATCCTTTCTGCCCACTTCCCCTTCCCCAATTACTATTCCCTTTGACTTCAGCTGCCTGAAACAGCCATGTCCAAGTTCTTCACCTCTATCCAA  2160
AGAACTTGATTTGCATGGATTTTGGATAAATCATTTCAGTATCATCTCCATCATATGCCTGACCCCTTGCTCCCTTCAATGCTAGAAAATCGAGTTGGCAAAATGGGGTTTGGGCCCCTC  2280
AGAGCCCTGCCCTGCACCCTTGTACAGTGTCTGTGCCATGGATTTCGTTTTTCTTGGGGTACTCTTGATGTGAAGATAATTTGCATATTCTATTGTATTATTTGGAGTTAGGTCCTCACT  2400
TGGGGGAAAAAAAAAAAAAAAAAAGCCAAGCAAACCAATGGTGATCCTCTATTTTGTGATGATGCTGTGACAATAAGTTTGAACCTTTTTTTTTGAAACAGCAGTCCCAGTATTCTCAGAGC  2520
ATGTGTCAGAGTGTTGTTCCGTTAACCTTTTTGTAAATACTGCTTGACCGTACTCTCACATGTGGCAAAATATGGTTTGGTTTTTCTTTTTTTTTTTTGAAAGTGTTTTTTCTTCGTCCT  2640
TTTGGTTTAAAAAGTTTCACGTCTTGGTGCCTTTTGTGTGATGCCCCTTGCTGATGGCTTGACATGTGCAATTGTGAGGGACATGCTCACCTCTAGCCTTAAGGGGGGCAGGGAGTGATG  2760
ATTTGGGGGAGGCTTTGGGAGCAAAATAAGGAAGAGGGCTGAGCTGAGCTTCGGTTCTCCAGAATGTAAGAAAACAAAATCTAAAACAAAATCTGAACTCTCAAAAGTCTATTTTTTTAA  2880
CTGAAAATGTAAATTTATAAATATATTCAGGAGTTGGAATGTTGTAGTTACCTACTGAGTAGGCGGCGATTTTTGTATGTTATGAACATGCAGTTCATTATTTTGTGGTTCTATTTTACT  3000
TTGTACTTGTGTTTGCTTAAACAAAGTGACTGTTTGGCTTATAAACACATTGAATGCGCTTTATTGCCCATGGGATATGTGGTGTATATCCTTCCAAAAAATTAAAACGAAAATAAAGTA  3120
GCTGCGATTGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

* To whom correspondence should be addressed

1: Oncogene Res. 1987 Sep-Oct;1(4):343-55. Related Articles, Links

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources
Order Documents
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

A novel early growth response gene rapidly induced by fibroblast, epithelial cell and lymphocyte mitogens.

Sukhatme VP, Kartha S, Toback FG, Taub R, Hoover RG, Tsai-Morris CH.

Department of Human Genetics, Howard Hughes Medical Institute, Philadelphia, Pennsylvania.

Mitogens evoke many alterations in gene expression in eukaryotic cells. Genes which are activated rapidly and transiently, that are evolutionarily conserved and whose induction is shared by diverse cell types when exposed to different growth stimuli are likely to be of critical importance in transducing mitogenic signals and regulating cellular proliferation. c-myc and c-fos are the only known genes fulfilling these criteria. We report on the molecular cloning of a novel early growth response (egr) gene which also satisfies these conditions. In response to serum, its 3.7 kb mRNA is induced dramatically in mouse fibroblasts reaching a peak level at about 30 minutes that is ten times higher than the maximal value attained by c-fos mRNA. This transcript is induced by the tumor promoter 12-O-tetradecanoyl-phorbol-13-acetate and is "superinduced" by serum and cycloheximide together. Importantly, the gene is highly induced by different mitogens in a wide array of cell types: insulin stimulated rat hepatoma cells, adenosine diphosphate treated monkey kidney epithelial cells, and phytohemagglutinin stimulated human peripheral blood lymphocytes. Given the many properties that this gene shares with c-myc and c-fos, it may play a key role in the control of cell growth and perhaps in oncogenesis.

PMID: 3130602 [PubMed - indexed for MEDLINE]

1: Cell. 1988 Apr 8;53(1):37-43.

Related Articles, Links

A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization.

Sukhatme VP, Cao XM, Chang LC, Tsai-Morris CH, Stamenkovich D, Ferreira PC, Cohen DR, Edwards SA, Shows TB, Curran T, et al.

Department of Medicine, Howard Hughes Medical Institute, University of Chicago, Illinois 60637.

Egr-1 is an early growth response gene that displays fos-like induction kinetics in fibroblasts, epithelial cells, and lymphocytes following mitogenic stimulation. Sequence analysis of murine Egr-1 cDNA predicts a protein with three DNA binding zinc fingers. The human EGR1 gene maps to chromosome 5 (bands 5q23-31). Egr-1 mRNA increases dramatically during cardiac and neural cell differentiation, and following membrane depolarization both in vitro and in vivo. Thus, Egr-1 and c-fos are often coregulated with strikingly similar kinetics. These results, in conjunction with the Egr-1 primary structure, suggest that Egr-1 may function as a transcriptional regulator in diverse biological processes.

PMID: 3127059 [PubMed - indexed for MEDLINE]

Display Abstract Show: 20 Sort Send to Text